مبادئ
الإقتصاد القياسي

Princilpes of econometrics

د. خالد محمد السواعي



مبادئ

الاقتصاد القياسي

Principles of Econometrics

د. خالد محمد السواعي

أستاذ الاقتصاد المساعد/ قسم الاقتصاد/ جامعة الزرقاء

رقم التصنيف : 330
المؤلف ومن هو في حكمه: خالد محمد السواعي
عنوان الكتاب : مبادئ الاقتصاد القياسي
رقم الإيداع : ر.إ. : 4991/ 10/ 2015
الواصفات : الاقتصاد القياسي/ الرياضيات
بيانات الناشر : دار الكتاب الثقافي

أعدت بيانات الفهرسة والتصنيف الأولية من قبل دائرة المكتبة الوطنية

1439 هـ - 2018 م

المحتويات



خالد محمد السواعي

أستاذ الاقتصاد المساعد في جامعة الزرقاء، بدأ دراسة الاقتصاد في جامعة اليرموك وحصل على البكالوريوس في عام ١٩٨٤، وتابع دراسته في الجامعة الأردنية واكمل فيها الماجستير (عام ٢٠٠٣) والدكتوراة (عام ٢٠١١)، وعمل في الجمارك الأردنية خلال الفترة ١٩٩١-٢٠١٤ إلى وصل لرتبة عميد جمارك.

له العديد من المؤلفات والأبحاث العلمية الأكاديمية المحكمة في عدة حقول: في اقتصاديات النمو والاقتصاد النقدي والتجارة الدولية والسياسة المالية والنقدية، وحصل على جائزة الباحثين الشباب في عام ٢٠٠٤، وعضو في الجمعية الأردنية للبحث العلمي، ومؤلف كتاب القياس الاقتصادي: المبادئ الأساسية وحالات تطبيقية، وكتاب موضوعات متقدمة في القياس الاقتصادي، ومدخل إلى القياس الاقتصادي، وأساسيات القياس الاقتصادي باستخدام EViews، وتحليل البيانات باستخدام SPSS، والتجارة الدولية: النظرية والتطبيق، والتجارة والتنمية.

مقدمة

أعددت هذا الكتاب ليكون رفيقاً لطالب الاقتصاد المبتدئ، فيبدأ بمقدمة توضح بعض المفاهيم الأساسية لتكوين النموذج الاقتصادي والحديث عن أنواع البيانات ثم يبدأ بشرح طريقة المربعات الصغرى العادية، ثم ينتقل إلى المشاكل القياسية الأكثر شيوعاً كعدم ثبات التباين، والاعتماد الخطي المتعدد، والارتباط الذاتي، وكيفية تشخيصها والتعامل معها، وينتقل إلى مناقشة نماذج المعادلات الآنيّة والمتغيّرات الوهمية. وهذا ما يناسب المرحلة الجامعية الأولى. ومن خلاله تستطيع:

- التعرف على نموذج الانحدار الكلاسيكي البسيط والمتعدد.

- التعرف على مشاكل نموذج الانحدار الكلاسيكي.

- التعرف على نماذج الانحدار غير الخطية.

- التعر.ف على بعض التقنيات كالمتغيّرات الوهمية، والمعادلات الآنية.

تم اختباره وتجريبة على طلاب الاقتصاد القياسي في جامعة الزرقاء لأكثر من فصل دراسي، وتم تصحيحه وتعديله ليكون مناسباً لطلاب هذا المساق.

والله ولي التوفيق

الدكتور خالد السواعي

عمّان في 2017/03/20



النموذج الاقتصادي
وطبيعة البيانات



الفصل الأول
تمهيد

تأتي دراسة الاقتصاد القياسي كجزء أساسي في دراسة الاقتصاد، وتزداد أهميته عند تقييم النظرية الاقتصادية ووضع الفرضيات، وقد تبين النظرية وجود علاقة بين متغيّرين أو أكثر، ويتم قياس العلاقة بين المتغيّرين ودراسة منهجية قياس العلاقات الاقتصادية باستخدام بيانات فعلية تسمى المنهجية باسم الاقتصاد القياسي **Econometrics**.

وتعني كلمة Econometrics "القياس (metrics في اليونانية) في الاقتصاد"، ويتضمن الاقتصاد القياسي جميع الأساليب الإحصائية والرياضية التي تستخدم في تحليل البيانات الاقتصادية، والهدف الأساسي من استخدام الأدوات الإحصائية والرياضية للبيانات الاقتصادية هو محاولة اثبات أو عدم اثبات فرضية أو نموذج اقتصادي.

1-1- نموذج الاقتصاد القياسي

النموذج الاقتصادي هو مجموعة من العلاقات الاقتصادية (متغيّرات) تصاغ بصيغ رياضية توضح سلوك هذه العلاقات لتبيّن عمل اقتصاد أو قطاع معيّن. وبذلك فهو إطار مبسط لتوضيح العلاقات المعقدة باستخدام الأساليب الرياضية في كثير من الأحيان.

نبدأ في البداية من نموذج أو نظرية اقتصادية، ومن هذه النظرية يتم صياغة النموذج القياسي المستخدم في الاختبار التجريبي، ثم يتم جمع البيانات اللازمة لاجراء الاختبار وتقدير النموذج.

بعد أن يتم تقدير النموذج نجري اختبارات التوصيف للتأكد من سلامة النموذج المستخدم، بالإضافة إلى الاختبارات التشخيصية لفحص أداء ودقة التقدير. فإذا أوضحت هذه الاختبارات أن النموذج مقبول ننتقل

إلى إجراء اختبار الفرضيات لفحص صلاحية التوقعات النظرية، وبالتالي القدرة على استخدام النموذج في إجراء التوقعات والتوصية بسياسات. أما إذا أوضحت الاختبارات التشخيصية والتوصيفية أن النموذج المستخدم غير مناسب، فعلى القياسيين Econometrician العودة إلى مرحلة صياغة النموذج وتعديله وإعادة جميع الإجراءات من مرحلة البدء. ويبحث هذا الكتاب هذه القضايا، ويزودك بجميع الأدوات الرياضية والتحليلية الأساسية لتمكينك من إجراء الاقتصاد القياسي النظري والتطبيقي.



شكل 1-1 مراحل التحليل القياسي

وعلية تتلخص خطوات بناء نموذج -بالرغم من وجود عدة مدارس فكرية بمنهجية الاقتصاد القياسي- على النحو التالي:

1. تحديد نظرية أو فرضية.
2. توصيف نموذج رياضي للنظرية.
3. توصيف نموذج إحصائي أو قياسي.
4. جمع البيانات.

5. تقدير معلمات النموذج الاقتصادي القياسي.
6. اختبار الفرضية.
7. التنبؤ أو التوقع.
8. استخدام النموذج لأغراض المراقبة أو السياسة.

لتوضيح الخطوات السابقة، دعونا النظر في نظرية معروفة هي نظرية الاستهلاك الكينزية.

1- تحديد نظرية أو فرضية

قال كينز: يزداد الاستهلاك في المتوسط بزيادة الدخل، لكن ليس بقدر زيادة الدخل، وافترض أن الميل الحدي للاستهلاك (MPC) هو معدل التغير في الاستهلاك بوحدة واحدة (دينار) إلى التغيّر في الدخل، ويكون أكبر من الصفر وأقل من 1.

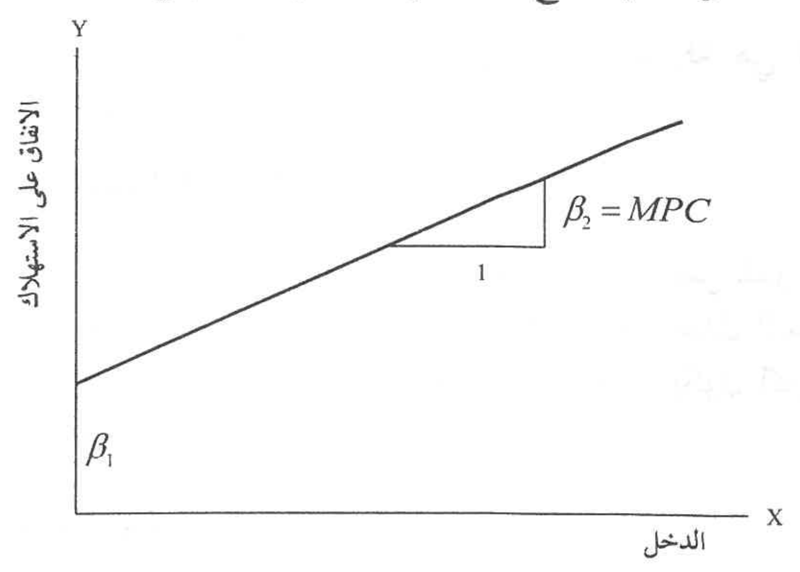
2- بناء نموذج رياضي

على الرغم من أن كينز يفترض وجود علاقة إيجابية بين الاستهلاك والدخل، فإنه لم يحدد بشكل دقيق دالة للعلاقة بينهما. وللتبسيط، قد يقترح الاقتصادي الرياضي الشكل التالي لدالة الاستهلاك الكينزية:

$$Y = \beta_1 + \beta_2 X \qquad 0 < \beta_2 < 1 \qquad (1.1)$$

حيث Y = الإنفاق الاستهلاكي و X = الدخل، و β_1 و β_2 تعرف باسم معالم النموذج، وهي على التوالي الحد الثابت ومعامل الانحدار. ويقيس معامل الميل β_2 الميل الحدي للاستهلاك. يبين الشكل (1-2) المعادلة (1.1)، وتنص هذه المعادلة على أن الاستهلاك يرتبط خطياً بالدخل، ويسمى النموذج الرياضي للعلاقة بين الاستهلاك والدخل بدالة الاستهلاك. والنموذج هو مجموعة معادلات رياضية؛ فإذا كان النموذج يتضمن معادلة واحدة فقط كما في المثال السابق، فهو يسمى بنموذج معادلة واحدة، في حين إذا كان يتضمن أكثر من معادلة، فهو يسمى بنموذج متعدد المعادلات.

يسمى المتغير الذي يظهر على الجانب الأيسر من المعادلة (1.1) من اشارة المساواة بالمتغير التابع، والمتغير(ات) على الجانب الأيمن تسمى متغير(ات) مستقلة أو تفسيرية. وهكذا في دالة الاستهلاك الكينزية: الاستهلاك (النفقات) هو المتغير التابع، والدخل هو المتغير التفسيري.



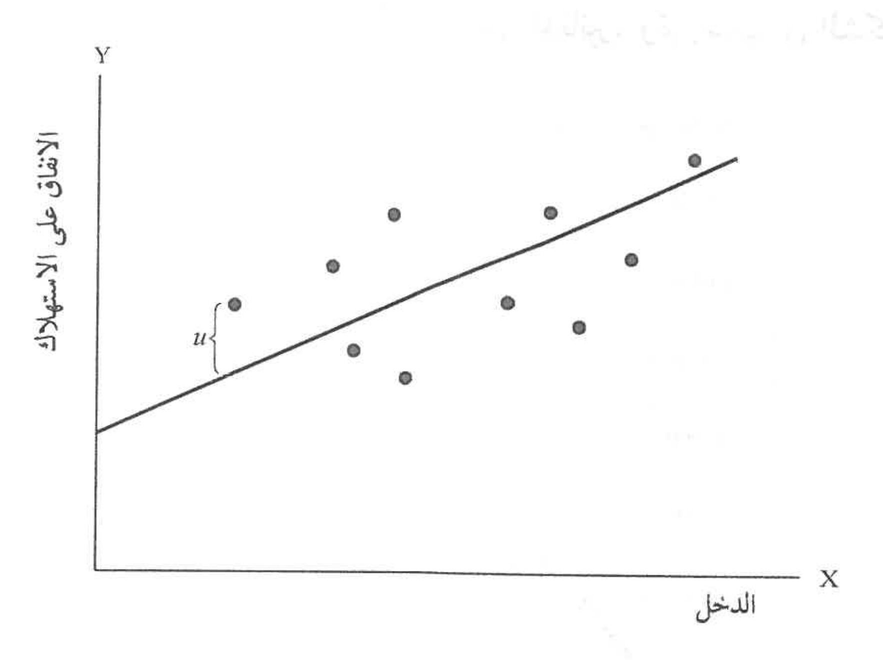
شكل رقم (1-2) دالة الاستهلاك الكينزية

3- بناء نموذج قياسي

إن العمل في ظل نموذج رياضي بحت لدالة الاستهلاك المعادلة (1.1) يفترض علاقة حتمية دقيقة بين الاستهلاك والدخل. إلا أن العلاقة بين المتغيرات الاقتصادية تكون غير دقيقة بشكل عام. وإذا حصلنا على بيانات عن نفقات الاستهلاك والدخل المتاح (بعد الضرائب) لعينة تمثل 500 أسرة أردنية مثلاً، واردنا رسم هذه البيانات على ورقة رسم بياني، نحدد الإنفاق الاستهلاكي على المحور الرأسي ونحدد الدخل المتاح على المحور الأفقي، ولا نتوقع أن جميع المشاهدات 500 تقع تماماً على خط مستقيم، لأن هناك متغيرات أخرى تؤثر على الإنفاق الاستهلاكي بالإضافة إلى الدخل، مثل حجم الأسرة، وأعمار أعضاء الأسرة، وديون الأسرة، وما إلى ذلك، ولها بعض التأثير على الاستهلاك.

فإذا كانت العلاقة بين المتغيرات الاقتصادية غير دقيقة، يعدل القياسيين دالة استهلاك المحددة (1.1) على النحو التالي:

$$Y = \beta_1 + \beta_2 X + u \qquad (1.2)$$

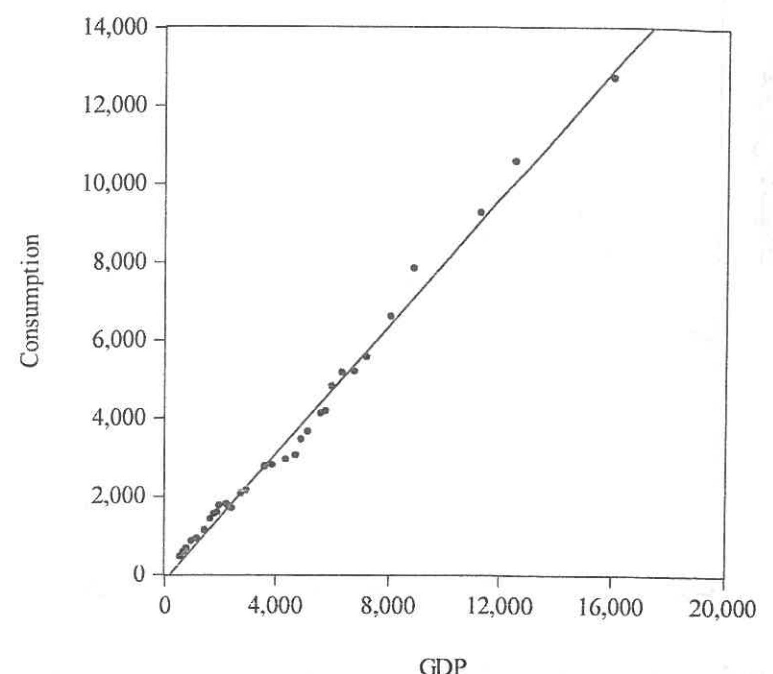


شكل (1-3): النموذج القياسي لدالة الاستهلاك الكينزية

حيث u تسمى بحد الاضطراب، أو حد الخطأ، وهو متغير عشوائي له خصائص احتمالية واضحة المعالم. ويمثل حد الاضطراب u جميع تلك العوامل التي تؤثر على الاستهلاك ولكن لا تؤخذ في الاعتبار بشكل صريح. وتمثل المعادلة (1.2) نموذج اقتصادي قياسي: ومن الناحية الفنية، هي مثال لنموذج الانحدار الخطي. وتفترض دالة الاستهلاك القياسية أن المتغير التابع Y (الاستهلاك) يرتبط خطياً بالمتغير التفسيري X (الدخل)، إلا أن العلاقة بينهما غير دقيقة، بل هي خاضعة للاختلافات الفردية. ويمكن تصور نموذج اقتصادي قياسي لدالة الاستهلاك كما في الشكل (1-3).

4- جمع البيانات

نحصل من تقدير النموذج الاقتصادي القياسى للمعادلة على القيم الرقمية للمعلمات β_1 و β_2 ونحتاج لبيانات تتصل بالاقتصاد الأردنى للفترة 1976-2007 على سبيل المثال، حيث يمثل المتغير Y مجموع الإنفاق الاستهلاكى الشخصى، ويمثل المتغير X الناتج المحلى الإجمالى بعد الضرائب (الدخل المتاح) وهو مقياس الدخل الإجمالي بملايين الدنانير، وتم رسمها في الشكل (4-1).



شكل رقم (4-1) العلاقة بين الإنفاق على الاستهلاك الخاص (Y) والدخل المتاح خلال الفترة 1976-2007 بملايين الدنانير الأردنية

5- تقدير النموذج القياسي

وبعد الحصول على البيانات يتم تقدير معلمات دالة الاستهلاك باستخدام الأسلوب الإحصائى لتحليل الانحدار الذي هو الأداة الرئيسية المستخدمة للحصول على التقديرات. وباستخدام هذه التقنية والبيانات التي تم جمعها نحصل على تقدير للمعلمتين β_1 و β_2، حيث بلغتا -273.4 و 0.7498 على التوالي، وبالتالي تكون دالة الاستهلاك المقدرة كما يلي:

$$\hat{Y}_i = -273.416 + 0.7498\ X_i \qquad (1.3)$$

وتشير القبعة ^ الظاهرة فوق Y على أنه تقدير. ويبين الشكل (1-4) تقدير دالة الاستهلاك (أي خط الانحدار).

Dependent Variable: Y
Method: Least Squares
Date: 06/05/11 Time: 12:56
Sample: 1976 2007
Included observations: 32

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-273.4160	104.3785	-2.619 67	0.0137
X	0.749767	0.018167	41.26971	0.0000

R-squared	0.982691	Mean dependent var	3247.813
Adjusted R-squared	0.982114	S.D. dependent var	2543.135
S.E. of regression	340.1161	Akaike info criterion	14.55691
Sum squared resid	3470370.	Schwarz criterion	14.64852
Log likelihood	-230.9106	Hannan-Quinn criter.	14.58728
F-statistic	1703.189	Durbin-Watson stat	0.534406
Prob(F-statistic)	0.000000		

وكما يبين الشكل (1-4)، فإن خط الانحدار المناسب للبيانات يكون عندما تكون نقاط البيانات قريبة جداً من خط الانحدار. ويبين هذا الشكل أن معامل الميل (MPC) حوالي 0.75، أي أن زيادة الدخل بمقدار دينار أردني واحد يؤدي إلى زيادة الإنفاق الاستهلاكي بالمتوسط بحوالي 75 قرشاً. ونقول في المتوسط لان العلاقة بين الاستهلاك والدخل ليست بالضبط، كما هو واضح من الشكل (1-4)؛ لأن كل نقاط البيانات لا تقع بالضبط على خط الانحدار. وبعبارة بسيطة نستطيع أن نقول أنه وفقاً لمعلوماتنا، يزداد الإنفاق الاستهلاكي في المتوسط بنحو 75 قرشاً عندما يزيد الدخل المتاح بمقدار دينار واحد.

6- اختبار الفرضية

على افتراض أن هذا النموذج هو تقريب جيد من الواقع إلى حد معقول، علينا وضع معايير مناسبة لمعرفة ما إذا كانت التقديرات التي تم الحصول عليها في المعادلة (1.3) تتفق مع التوقعات النظرية التي يتم اختبارها. ووفقاً لميلتون فريدمان، إذا لم نستطيع التحقق من النظرية أو الفرضية بالأدلة التجريبية فإنها لن تكون مقبولة.

كما لوحظ في وقت سابق، توقَّع كينز أن يكون الميل الحدي للاستهلاك موجباً وأقل من 1. ووجدنا في مثالنا أن الميل الحدي للاستهلاك نحو 0.75، ولكن قبل أن تقبل هذه النتيجة تأكيداً لنظرية الاستهلاك الكينزية، يجب أن نتساءل ما إذا كان هذا التقدير كافٍ لإقناعنا بأن حدوثه ليس صدفة؛ وبعبارة أخرى، فإن القيمة 0.75 إحصائياً هي أقل من 1 وتدعم نظرية كينز. ويستند هذا التأكيد (أو دحض النظرية الاقتصادية) على أساس أدلة من النظرية الإحصائية المعروفة باسم الاستدلال الإحصائي (اختبار الفرضيات).

7- التوقع أو التنبؤ

إذا كان النموذج المختار لا يدحض الفرضية أو النظرية قيد النظر، سيتم استخدامه للتنبؤ بالقيم المستقبلية للمتغير التابع، أو توقع المتغير Y على أساس القيمة المعروفة أو القيمة المستقبلية المتوقعة للمتغير التفسيري X. ولتوضيح ذلك، نفترض أننا نريد التنبؤ بمتوسط الإنفاق الاستهلاكي لعام 2008. فإذا توقعنا بأن قيمة الناتج المحلي الإجمالي في عام 2008 حوالي 18144 مليون دينار أردني، وتم تعويض هذا الرقم بدلاً من متغيّر الناتج المحلي الإجمالي على الجانب الأيمن من المعادلة (1.3) سنحصل على:

$$\hat{Y}_{2008} = -273.416 + 0.7498\,(18144)$$
$$= 13331 \qquad (1.4)$$

أي حوالي 13331 مليون دينار، وبناءً على قيمة الناتج المحلي الإجمالي، سيكون توقع متوسط نفقات الاستهلاك حوالي 13331 مليون دينار، علماً

بأن القيمة الفعلية لنفقات الاستهلاك في عام 2008 كانت تساوي 12726.4 مليون دينار. وأوضح النموذج المقدر (1.3) أن نفقات الاستهلاك الفعلي أكبر من المتوقع بنحو 605 مليون دينار. ويمكننا القول بأن خطأ التوقع حوالي 605 مليون دينار، وهي عبارة عن 3.3٪ من قيمة الناتج المحلي الإجمالي الفعلي لعام 2008، وعندما نناقش نموذج الانحدار الخطي في الفصول اللاحقة، سنحاول معرفة ما إذا كان مثل هذا الخطأ "صغيراً" أو "كبيراً" ولكن المهم الآن هو ملاحظة أن هذه التوقعات لا تخلو من أخطاء نظراً للطبيعة الإحصائية لتحليلنا.

وهناك استخدام آخر للنموذج المقدر (1.3). لنفترض أن مجلس الوزراء قرر إجراء تخفيض على ضريبة الدخل، ماذا سيكون تأثير هذه السياسة على الدخل وبالتالي على الإنفاق الاستهلاكي وبالتالي على التشغيل؟

نفترض أن نتيجة هذا التغيير في السياسة المقترحة، تكون زيادة النفقات الاستثمارية. ماذا سيكون التأثير على الاقتصاد؟ وكما تبين نظرية الاقتصاد الكلي، أن التغير في الدخل يتبع التغير في الإنفاق الاستثماري المعطى بمضاعف الدخل الذي يعرَّف كما يلي:

$$M = \frac{1}{1 - MPC} \qquad (1.5)$$

إذا استخدمنا الميل الحدي للاستهلاك الذي تم الحصول عليه 0.75 في المعادلة (1.3)، يصبح المضاعف M = 4. وهذا يعني أن زيادة (نقصان) الاستثمار بمقدار دينار واحد سوف يؤدي في نهاية المطاف إلى زيادة (نقص) الدخل بأربعة أضعاف، علماً بأن العملية تتطلب بعض الوقت ليعمل المضاعف.

القيمة الحرجة في هذا الحساب هي الميل الحدي للاستهلاك، ويعتمد المضاعف على ذلك. ويمكن الحصول على هذا التقدير من الميل الحدي للاستهلاك من نماذج الانحدار مثل المعادلة (1.3). وهكذا، فإن التقدير الكمي للميل الحدي للاستهلاك يوفر معلومات قيّمة لأغراض السياسة.

ومن معرفة الميل الحدي للاستهلاك يمكن للمرء التنبؤ بمسار الإيرادات والنفقات والاستهلاك والعمالة المستقبلية في أعقاب حدوث تغيير في السياسات المالية للحكومة.

8- استخدام النموذج في السياسة الاقتصادية

لنفترض أنه لدينا تقدير لدالة الاستهلاك المعطاة، ولنفترض أن الحكومة تعتقد أن زيادة الإنفاق الاستهلاكي بحوالي 13000 مليون دينار تُبقي معدل البطالة عند مستواه في عام (2008) البالغ نحو 12.7٪. فما هو مستوى الدخل الذي يضمن المبلغ المستهدف من الإنفاق الاستهلاكي؟ إذا كانت النتائج الواردة في الانحدار (1-3) تبدو معقولة، وبعملية حسابية بسيطة نحصل على:

$$13000 = -273.416 + 0.7498\ X_i \qquad (1.6)$$

تكون قيمة X = 17703 مليون تقريباً؛ أي أنه عند مستوى الدخل 17703 مليون دينار، وميل حدي للاستهلاك حوالي 0.75، سوف ينتج نفقات حوالي 13000 مليون دينار. كما تشير هذه الحسابات، أنه يمكن استخدام النموذج المقدر بهدف المراقبة أو السياسة. وبمزيج مناسب من السياسات المالية والنقدية، يمكن للحكومة التعامل بالمتغير X لإنتاج المستوى المنشود لمتغير الهدف Y.

1-2- أنواع البيانات الاقتصادية

تأخذ البيانات الاقتصادية عدة أشكال هي:

1-2-1 البيانات المقطعية Cross-Sectional Data

تتكون مجموعة البيانات المقطعية من عينة الأفراد، أو القطاع العائلي، أو الشركات، أو الدول، او المناطق، أو المدن، أو أي نوع من الوحدات في نقطة محددة من الزمن. وفي بعض الحالات، لا تتماثل الفترة الزمنية للبيانات بالضبط؛ مثل مسح بيانات العائلات المختلفة خلال أيام مختلفة من الشهر،

وفي هذه الحالة يتم اهمال فروق التوقيت في جمع البيانات وتسمى البيانات التي يتم جمعها بيانات مقطعية.

إن استخدام البيانات المقطعية واسع النطاق في الاقتصاد والعلوم الاجتماعية الأخرى. ويساهم تحليل البيانات المقطعية بالاقتصاد الجزئي التطبيقي في اقتصاديات العمل، والمالية العامة للدولة، والاقتصاد الإداري، والاقتصاد الطبي، وبعض الحقول في الاقتصاد الجزئي مثل بيانات الأفراد، والعائلات، والشركات، والمدن، والمناطق في نقطة من الزمن، وتستخدم تلك الحالات لاختبار فرضيات الاقتصاد الجزئي، وتقييم السياسات الاقتصادية.

يمكن تمثيل البيانات المقطعية المستخدمة في التحليل الاقتصادي القياسي على شكل الجدول (1-1) الذي يحتوي على بيانات مقطعية ل 526 شخص عملوا في عام 1976. وتشمل المتغيرات الأجور (دينار لكل ساعة)، وسنوات التعليم، وسنوات الخبرة، ومؤشر النوع (ذكر، أنثى)، الحالة الاجتماعية. ويأخذ المتغيران السابقان قيم ثنائية (صفر، واحد) للإشارة إلى الميزات النوعية للفرد (الشخص هل هو أنثى أم ذكر؛ الشخص متزوج أم لا). ولا يوجد أي اعتبار لترتيب البيانات المقطعية (لا تشبه بيانات السلاسل الزمنية).

جدول (1-1) بيانات مقطعية عن الأجور والخصائص الفردية الأخرى

المشاهدة	الأجور	التعليم	الخبرة	النوع	الحالة الاجتماعية
1	3.10	11	2	1	0
2	3.24	12	22	1	1
3	3.00	11	2	0	0
4	6.00	8	44	0	1
5	5.30	12	7	0	1
⋮	⋮	⋮	⋮	⋮	⋮
525	11.56	16	5	0	1
526	3.50	14	5	1	0

2-2-1 بيانات السلاسل الزمنية Time Series Data

تتكون بيانات السلاسل الزمنية من مشاهدات متغيّر واحد أو أكثر خلال فترة من الزمن، منظمة بترتيب تسلسلي زمني مثل: سنوي، نصف سنوي، فصلي، شهري، أسبوعي، يومي، كل ساعة. ومن الأمثلة على بيانات السلاسل الزمنية أسعار الأسهم، والناتج المحلي الإجماليGDP، وعرض النقد، وغيرها.

ومن الأمثلة على السلاسل الزمنية الجدول (1-2) الذي يتضمن سلسلة قيمة المستوردات الخاضعة للرسوم الجمركية وقيمة الرسوم الجمركية بالمليون دينار، وكما يلي:

جدول (2-1)

المستوردات الخاضعة للرسوم وقيمة الرسوم الجمركية المتحققة عليها (مليون دينار)

السنة	المستوردات الخاضعة للرسوم الجمركية	الرسوم الجمركية
2003	857.1	162.6
2004	1100.6	211.6
2005	1194.6	254.7
2006	1357.0	280.3
2007	1467.2	279.0
2008	1390.5	254.5
2009	1174.5	237.5
2010	1066.0	235.5
2011	1289.5	255.9
2012	1425.1	259.5
2013	1563.7	298.8
2014	1412.3	299.1
2015	1355.2	298.9

المصدر: دائرة الجمارك الأردنية، التقارير السنوية.

وطبيعة البيانات الزمنية تجعل من تحليلها صعوبة أكبر من تحليل البيانات المقطعية، وتعتمد البيانات الاقتصادية على الزمن، وهذا يعني أن أغلب بيانات السلاسل الزمنية الاقتصادية ترتبط بتاريخها الماضي، وتطبق أغلب الإجراءات القياسية على البيانات المقطعية وعلى بيانات السلاسل الزمنية. إلا أننا نحتاج في حالة بيانات السلاسل الزمنية إلى إجراءات لتحديد النموذج القياسي المناسب، كما أن البيانات الاقتصادية الزمنية تتضمن اتجاهاً زمنياً يقودنا إلى أساليب قياسية جديدة.

3-2-1 بيانات السلاسل المقطعية – الزمنية Panel Data

تتكون بيانات Panel من سلاسل زمنية لكل طرف مقطعي في مجموعة البيانات: مثل المبيعات وعدد العاملين لخمسين شركة خلال فترة خمس سنوات. ويمكن جمع بيانات Panel على الأساس الجغرافي؛ مثل الناتج المحلي الإجمالي وعرض النقد لعشرين دولة لفترة 20 سنة.

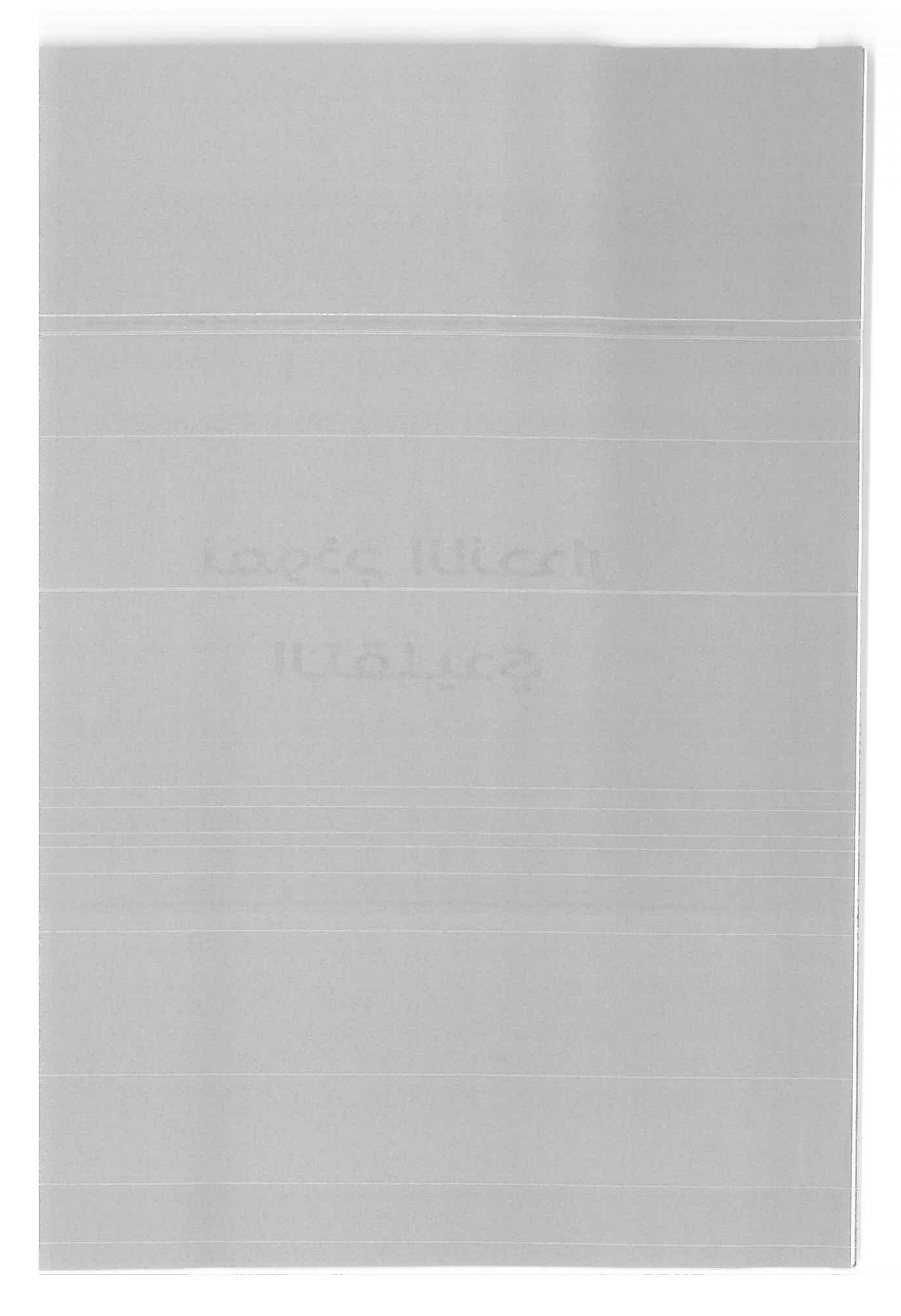
مثلاً تتضمن بيانات عشرة دول خلال الفترة 1990-1999 للمتغير $Y = GDP$ قيمة الناتج المحلي الإجمالي GDP لكل للدولة في عام 1990 يتبعها GDP لنفس الدولة في عام 1991 و ... وهكذا، خلال الفترة T سنة

جدول رقم (3-1) بيانات سلاسل زمنية مقطعية مجمعة: بيانات اقتصادية لبعض الدول

Country	code	year	gdp	save	pop
Algeria	2	1990	2.3	27.5	2.5
Algeria	2	1991	-3.7	36.7	2.4
Algeria	2	1992	-3.6	32.4	2.4
Algeria	2	1993	-0.8	27.8	2.3
Algeria	2	1994	-4.4	27.0	2.2
Algeria	2	1995	-3.3	28.4	2.2
Algeria	2	1996	1.6	31.4	2.2
Algeria	2	1997	1.6	32.2	2.2
Algeria	2	1998	-1.0	27.1	2.1
Algeria	2	1999	1.4	31.7	2.1

Angola	3	1990	-2.4	29.7	3.1
Angola	3	1991	-3.3	16.2	3.9
Angola	3	1992	-3.1	1.7	3.6
↓		↓	↓	↓	↓
Angola	3	1998	3.5	32.5	2.9
Angola	3	1999	-2.9		2.9
Argentina	4	1990	-8.8	19.7	1.3
Argentina	4	1991	-3.7	16.2	1.3
↓		↓	↓	↓	↓
Argentina	4	1999	-4.7	17.2	1.2
Bahrain	9	1990	-2.1	42.4	2.8
Bahrain	9	1991	-1.5	35.7	1.0
Bahrain	9	1992	3.5	33.0	2.1
Bahrain	9	1993	5.5	35.9	3.4
Bahrain	9	1994	4.7	31.9	3.7
Bahrain	9	1995	-1.2	36.9	3.5
Bahrain	9	1996	-1.3	40.1	3.7
Bahrain	9	1997	-0.6	42.1	3.4
Bahrain	9	1998	-0.3		3.6
Bahrain	9	1999			3.3

نموذج الانحدار التقليدي



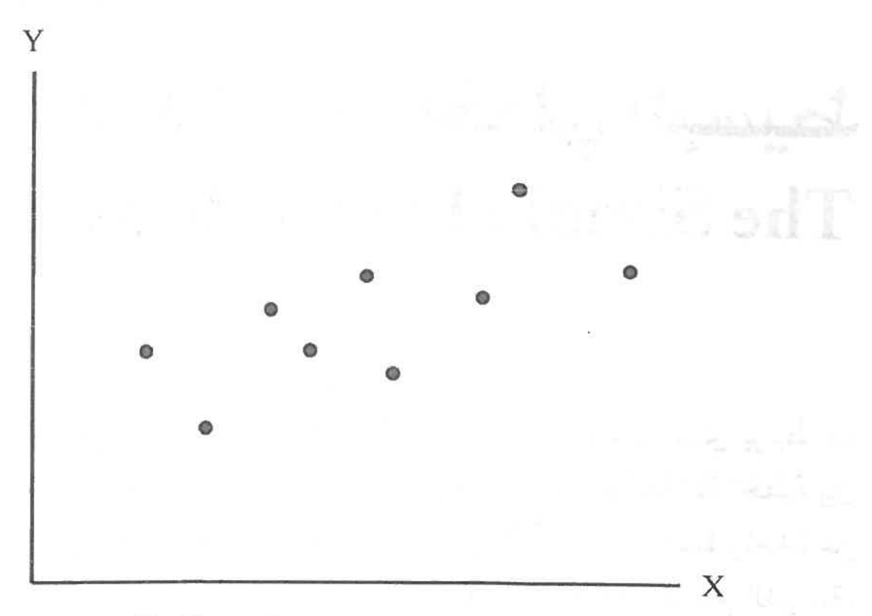
الفصل الثاني
نموذج الانحدار الخطي البسيط
The Simple Linear Regression

يستعرض هذا الفصل نموذج الانحدار الخطي البسيط الذي يرتبط فيه متغيّر واحد (X) بمتغيّر آخر (Y)، ويفترض هذا النموذج العلاقة الخطية بين X و Y، ويمثل ميل الخط الذي يربط بين X و Y أثر تغيّر وحدة واحدة من X على Y. كما أن خصائص توزيع وسط Y غير معروفة لمجتمع الدراسة، وكذلك ميل العلاقة الخطية بين X و Y غير معروفة للمجتمع، والمشكلة القياسية هي تقدير الميل لكي نقدّر الأثر على Y عندما يتغيّر X بوحدة واحدة باستخدام بيانات عينة هذان المتغيّران.

يصف هذا الفصل منهجية تقدير هذا الميل باستخدام عينة عشوائية من بيانات X و Y؛ مثل استخدام بيانات الدخل الفردي واستهلاك عائلات مختلفة ومنفصلة، وسنرى كيف نقدر الأثر المتوقع لاستهلاك العائلات عندما ينخفض دخل العائلة الواحدة، ويمكن تقدير الميل Slope والحد الثابت Intercept لخط العلاقة بين X و Y بمنهجية تسمى المربعات الصغرى العادية (OLS) Ordenary Least Squered.

2-1- نموذج الانحدار الخطي THE LINEAR REGRESSION MODEL

افرض أن الباحث لديه فكرة تُبني على علاقة بين متغيّرين هما Y و X، وتوضح النظرية الاقتصادية أن زيادة X ستؤدي إلى زيادة Y، بداية نختبر فيما إذا كانت هناك علاقة بين المتغيرين برسم انتشارها، وافرض أن الناتج يبيّنه الشكل (2-1) التالي:



شكل (2-1) انتشار بيانات المتغيرين Y و X

يظهر هذا الشكل علاقة خطية موجبة بين Y و X؛ تعني أن أي زيادة في X تؤدي إلى زيادة في Y، ويمكن توصيف العلاقة بينهما بخط مستقيم، ومن الممكن رسمها باليد بشكل خط يظهر توزيع البيانات، إلا أن توزيع المقطع وميل الخط بالعين المجردة يبدو شاقاً وغير دقيق، وبناءً عليه سنكون مهتمين بتحديد العلاقة بالضبط بمعادلة نستطيع تقديرها باستخدام إجراء محدد، ومن الممكن استخدام معادلة الخط المستقيم التالية:

$$Y_i = \beta_1 + \beta_2 X_i \quad (2.1)$$

للحصول على أفضل توزيع خطي للبيانات، يسعى الباحث الحصول على قيم المعاملات **parameters** أو المعلمات **coefficients** β_1 و β_2 لتقدير خط ينطبق قدر الامكان على جميع نقاط البيانات.

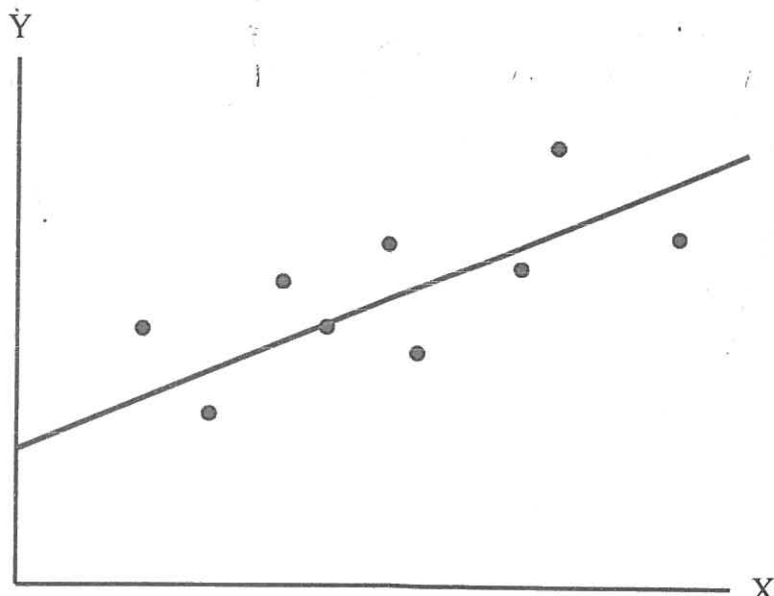
هذه المعادلة $Y_i = \beta_1 + \beta_2 X_i$ تمثل بالضبط خط واحد، وعلى فرض أنها كانت مناسبة وحسبت قيم β_1 و β_2، وإذا توفَّر لدينا قيم X

نستطيع تحديد ما ستكون عليه قيم Y؛ أي أن قيمة أي متغيّر تعطي قيمة المتغيّر الآخر.

هذا النموذج ليس واقعياً، لأنه يجب أن يتوافق احصائياً مع النموذج المقدر بالضبط، وأن جميع النقاط تقع على الخط المستقيم بالضبط، ولجعل هذا النموذج أكثر واقعية سنضيف للمعادله حد الخطأ العشوائي الذي يشار إليه بالرمز u_i كما يلي:

$$Y_i = \beta_1 + \beta_2 X_i + u_i \qquad (2.2)$$

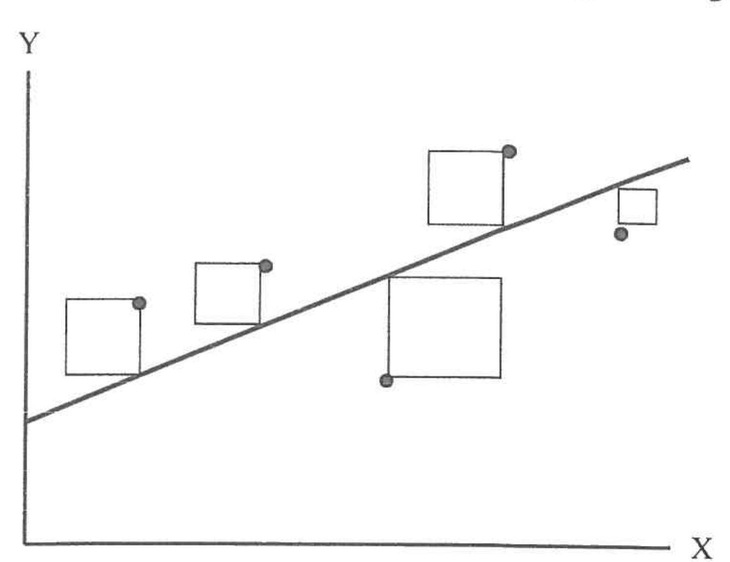
كيف تتحدّد قيم β_1 و β_2 المناسبة؟ يتم اختيار β_1 و β_2 بحيث يتمّ تخفيض المسافة بين نقاط البيانات والخط المقدّر، وأن تلتصق البيانات بالخط المقدَّر قدر الامكان، لذا يتم اختيار المعلمات التي تخفّض المسافة العمودية بين نقاط البيانات والخط المقدَّر، كما يوضحة الشكل (2-2) التالي:



شكل (2-2) انتشار بيانات المتغيّرين مع أفضل خط مختار بالنظر

لكن هذا الأسلوب غير مناسب ويبدو غير دقيق، ومن أشهر الطرق المستخدمة لتقدير الخط المناسب طريقة تسمى بالمربعات الصغرى العادية Ordinary least squars (OLS)، ويعتبر هذا الأسلوب العمود الفقري لتقدير النموذج القياسي الذي سنشرحه بالتفصيل في هذا الفصل.

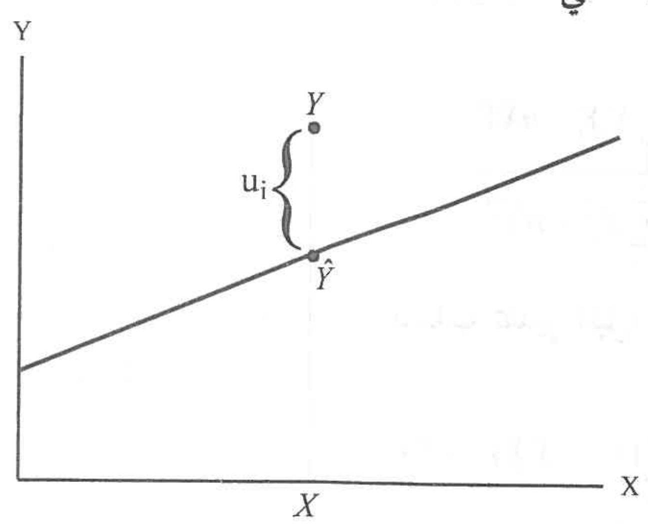
افرض أنه يوجد لدينا عينة مكوّنة من 5 مشاهدات، ويستلزم أسلوب المربعات الصغرى العادية OLS أخذ مسافة عمودية من كل نقطة إلى الخط وتربيعها وتخفيض مجموع مربع المساحة الكلي (من هنا جاءت المربعات الصغرى) كما يظهره الشكل (2-3)، وهذا يكافئ مجموع المساحات المربعة المرسومة من النقاط إلى الخط.



شكل (2-3) أسلوب المربعات الصغرى في تقدير أنسب خط يُخفض مجموع المربعات

تشير Y_i إلى نقطة البيانات الفعلية للمشاهدة i و $\hat{Y}_i$ القيمة المقدرة من معادلة خط الانحدار؛ وبعبارة أخرى، إذا كان لديك قيمة X_i للمشاهدة i، تكون $\hat{Y}_i$ القيمة التي تنبأ بها النموذج للمتغيّر Y_i، كما تشير الإشارة ^ فوق المتغيّر أو القيمة المقدّرة للمعامل في النموذج. أخيراً تشير

u_i إلى البواقي التي هي الفرق بين القيمة الفعلية Y_i والقيمة المقدّرة بالنموذج لهذه النقطة؛ أي $(Y_i - \hat{Y}_i)$، وتظهر هذه كمشاهدة واحدة في الشكل (2-4) التالي:



شكل (2-4) رسم مشاهدة واحدة مع أفضل خط مقدّر والبواقي والقيمة المقدّرة

يتم تخفيض مجموع مربع المساحات التي يعبّر عنها $(\hat{u}_1^2 + \hat{u}_2^2 + \cdots + \hat{u}_n^2)$ أو تخفيض: $\sum \hat{u}_i^2$ *Minimizing*، ويسمى هذا المجموع بمجموع مربعات البواقي **Residual Sum of Squares RSS** أو **Sum of Squared Residuals SSR**، لكن ما هي $\hat{u}_i$؟ هي الفرق بين النقاط الفعلية والخط المقدر $Y_i - \hat{Y}_i$،

وعليه فإن: *Minimizing* $\sum_{i=1}^{n} \hat{u}_i^2$ يساوي *Minimizing* $\sum (Y_i - \hat{Y}_i)^2$،

وعلى افتراض أن b_1 و b_2 تشير إلى تقدير قيم β_1 و β_2 تعطى معادلة الخط المقدر كما يلي: $\hat{Y}_i = b_1 + b_2 X_i$. وبما أن SSR تشير إلى مجموع مربع البواقي أو الأخطاء العشوائية، فإن:

$$SSR = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{n} (Y_i - b_1 - b_2 X_i)^2 \qquad (2.3)$$

ستُخفّض SSR بالنسبة للمعلمات b_1 و b_2 لايجاد قيم β_1 و β_2 التي تُخفّض مجموع مربع البواقي للحصول على خط يتطابق مع البيانات، وبالتالي يشتق SSR بالنسبة b_1 و b_2 وتُساوى بالصفر. وتعطى المعلمات المقدرة للميل والمقطع كما يلي:

$$b_2 = \frac{\sum_{i=1}^{n} X_i Y_i - n\bar{X}\bar{Y}}{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2} \qquad (2.4)$$

والمعادلة أعلاه سهلة الاستخدام لحساب تقدير الميل؛ إلا أن هذه الصيغة يمكن كتابتها كما يلي:

$$b_2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n} (X_i - \bar{X})^2} \qquad (2.5)$$

كما أن صيغة تقدير الحد الثابت (المقطع):

$$b_1 = \bar{Y} - b_2 \bar{X} \qquad (2.6)$$

لماذا حد الخطأ؟

هناك عدة أسباب لاضافة حد الخطأ إلى المعادلة القياسية:

1 - إغفال بعض المتغيّرات التفسيرية: العلاقة بين Y و X بسيطة، وفي الواقع يوجد عوامل أخرى تؤثر على Y غير موجودة في المعادلة (2.1)، وسيؤدي تأثيرها إلى وقوع نقاط خارج الخط، وقد ترغب باضافة العديد من المتغيّرات في معادلة الانحدار، لكنك لا تستطيع إضافتها لأنك لا ترغب قياسها، وبذلك تشترك جميع العوامل الأخرى في تكوين حد الخطأ.

2 - المتغيّرات المجمّعة: في بعض الحالات يتم تجميع بيانات لمجموعة أفراد من أجل دراسة علاقات خاصة بالاقتصاد الجزئي كدالة الاستهلاك الكلي؛ حيث يتم تلخيص مجموعة قرارات الإنفاق الفردي، وعلى الأرجح يكون للعلاقات الفردية معلمات مختلفة، فأي محاولة لتقدير علاقة الإنفاق الكلي على الدخل الكلي تكون تقريبية، ويتم تفسير التفاوت بحد الخطأ.

3 - سوء توصيف النموذج: قد يكون هيكل النموذج سيء التوصيف، مثلاً إذا أشارت العلاقة إلى بيانات سلسلة زمنية، فقد لا تعتمد قيم Y على قيم X الفعلية؛ إنما تعتمد على قيم متوقعة في فترة سابقة، فإذا كانت القيم المتوقعة والفعلية مرتبطة ارتباطاً وثيقاً ستبدو العلاقة بين Y و X تقريبية، وسيلتقط حد الخطأ الفرق بينهما.

4 - سوء توصيف الدالة: قد تكون دالة العلاقة بين Y و X سيئة التوصيف رياضياً، فمثلاً قد تكون العلاقة الصحيحة علاقة غير خطية، ولتجنب هذه المشكلة نستخدم التوصيف الرياضي المناسب، إلا أن تطوير التوصيف قد يكون تقريبي على الأرجح، ويساهم الفرق في حد الخطأ.

5 - أخطاء القياس: إذا كان قياس متغيّر واحد أو أكثر في العلاقة عرضة للخطأ؛ فإن القيم المشاهدة لا تظهر مطابقة للعلاقة الفعلية ويساهم الفرق في حد الخطأ.

وعليه فإن حد الخطأ (الاضطراب) هو حاصل تجميع تلك العناصر، وبالتأكيد إذا أخذنا قياس أثر X على Y سيكون الأثر مناسباً إذا لم يوجد حد الخطأ، ولولا وجودها ستتطابق النقاط في الشكل (2-2) مع الخط، وعليك ملاحظة أن كل تغيّر في Y (من مشاهدة إلى مشاهدة) هو نتيجة التغيّر في X، وتستطيع حساب β_1 و β_2 بالضبط. وفي الحقيقة فإن جزءًا من التغيّر في Y يكون نتيجة التغيّر في u، وهذا يجعل الأمر أكثر صعوبة، ولهذا السبب توصف u في بعض الأحيان بالإزعاجات *noise*.

2-2- اشتقاق معاملات انحدار المربعات الصغرى بمتغيّر تفسيري واحد

سبق وأن أوضحنا أنه يتم تخفيض مجموع مربعات البواقي (حدود الخطأ العشوائي) $(\hat{u}_1^2+\hat{u}_2^2+\cdots+\hat{u}_n^2)$ أو $\sum_{i=1}^{n}\hat{u}_i^2$ ***Minimizing*** للحصول على خط أقرب للبيانات؛ ويسمى SSR ويُعبِّر عنه كما يلي:

$$SSR=\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2=\sum_{i=1}^{n}(Y_i-b_1-b_2\ X_i)^2 \qquad (2.7)$$

من الضروري تخفيض مجموع مربعات الأخطاء SSR للمعلمات b_1 و b_2 لايجاد قيم β_1 و β_2 التي تخفّض مجموع مربع البواقي للحصول على خط يتطابق مع البيانات، وبالتالي يشتق SSR بالنسبة b_1 و b_2 وتُساوى بالصفر. وتُعطى المعلمات المقدرة للميل والمقطع كما يلي:

نأخذ الشرط الأول $\partial SSR/\partial b_1=0$ و $\partial SSR/\partial b_2=0$ ونحصل على المعادلات التالية:

$$\frac{\partial SSR}{\partial b_1}=-2\sum_{i=1}^{n}(Y_i-b_1-b_2\ X_i)=0 \qquad (2.8)$$

$$\frac{\partial SSR}{\partial b_2}=-2\sum_{i=1}^{n}X_i\ (Y_i-b_1-b_2\ X_i)=0 \qquad (2.9)$$

تسمى هاتان المعادلتان بالمعادلات الطبيعية **Normal equations** لمعاملات الانحدار، والخطوة التالية اعادة ترتيب (2.8) و (2.9) للحصول على صيغة b_1 و b_2، ومن (2.8):

$$\sum_{i=1}^{n}(Y_i-b_1-b_2\ X_i)=0 \qquad (2.10)$$

يتم توزيع الأقواس ليمتد المجموع من 1 إلى n لتكون n مع b_1.

$$\sum_{i=1}^{n} Y_i - N b_1 - b_2 \sum_{i=1}^{n} X_i = 0 \qquad (2.11)$$

وبما أن $\sum_{i=1}^{n} Y_i = N\bar{Y}$ و $\sum_{i=1}^{n} X_i = N\bar{X}$ نستطيع كتابة (2.11) كما يلي:

$$N\bar{Y} - N b_1 - N b_2 \bar{X} = 0 \qquad (2.12)$$

أو،

$$\bar{Y} - b_1 - b_2 \bar{X} = 0 \qquad (2.13)$$

ومن (2.13) نحصل على:

$$\boldsymbol{b_1 = \bar{Y} - b_2 \bar{X}} \qquad \mathbf{(2.14)}$$

ومن (2.9) نحصل على:

$$\sum_{i=1}^{n} X_i (Y_i - b_1 - b_2 X_i) = 0 \qquad (2.15)$$

عوض b_1 من (2.14) في (2.15) نحصل على:

$$\sum_{i=1}^{n} X_i (Y_i - \bar{Y} + b_2 \bar{X} - b_2 X_i) = 0 \qquad (2.16)$$

$$\sum_{i=1}^{n} X_i Y_i - \bar{Y} \sum_{i=1}^{n} X_i + b_2 \bar{X} \sum_{i=1}^{n} X_i - b_2 \sum_{i=1}^{n} X_i^2 = 0 \qquad (2.17)$$

$$\sum_{i=1}^{n} X_i Y_i - N\bar{X}\bar{Y} + b_2 N \bar{X}^2 - b_2 \sum_{i=1}^{n} X_i^2 = 0 \qquad (2.18)$$

نعيد ترتيبها بالنسبة للمعلمة b_2

$$b_2\left(N\overline{X}^2-\sum_{i=1}^{n}X_i^2\right)=N\overline{X}\overline{Y}-\sum_{i=1}^{n}X_iY_i \qquad (2.19)$$

اقسم كلا الجانبين على $N\overline{X}^2-\sum_{i=1}^{n}X_i^2$ نحصل على:

$$b_2=\frac{\sum_{i=1}^{n}X_iY_i-N\,\overline{X}\overline{Y}}{\sum_{i=1}^{n}X_i^2-N\overline{X}^2} \qquad (2.20)$$

وهناك شكل بديل لهذا المعامل نفضله، وهو كما يلي:

$$b_2=\frac{\sum\left(X_i-\overline{X}\right)\left(Y_i-\overline{Y}\right)}{\sum\left(X_i-\overline{X}\right)^2} \qquad (2.21)$$

ولإثبات هذا،

$$\begin{aligned}\sum\left(X_i-\overline{X}\right)\left(Y_i-\overline{Y}\right)&=\sum_{i=1}^{n}X_iY_i-\sum_{i=1}^{n}X_i\overline{Y}-\sum_{i=1}^{n}\overline{X}Y_i+\sum_{i=1}^{n}\overline{X}\overline{Y}\\&=\sum_{i=1}^{n}X_iY_i-\overline{Y}\sum_{i=1}^{n}X_i-\overline{X}\sum_{i=1}^{n}Y_i+N\,\overline{X}\overline{Y}\\&=\sum_{i=1}^{n}X_iY_i-\overline{Y}\left(N\overline{X}\right)-\overline{X}\left(N\overline{Y}\right)+N\,\overline{X}\overline{Y}\\&=\sum_{i=1}^{n}X_iY_i-N\,\overline{X}\overline{Y}\end{aligned}$$

وبالمثل،

$$\sum\left(X_i-\overline{X}\right)^2=\sum_{i=1}^{n}X_i^2-N\overline{X}^2$$



مثال

لديك الآن قيم لمتغيرين هما Y و X يمثلان المتغيّر التابع والمتغيّر المستقل على التوالي. والمطلوب تقدير معاملات معادلة الانحدار الخطي البسيط، واكتبها على صيغة معادلة خطية بسيطة.

البيانات الخام			العمليات الحسابية			
i	X	Y	$(X-\bar{X})$	$(Y-\bar{Y})$	$(X-\bar{X})(Y-\bar{Y})$	$(X-\bar{X})^2$
1	20	30	-2	-15	30	4
2	40	60	18	15	270	324
3	20	40	-2	-5	10	4
4	30	60	8	15	120	64
5	10	30	-12	-15	180	144
6	10	40	-12	-5	60	144
7	20	40	-2	-5	10	4
8	20	50	-2	5	-10	4
9	20	30	-2	-15	30	4
10	30	70	8	25	200	64
$\sum$	**220**	**450**	**0**	**0**	**900**	**760**
	$\bar{X}$	$\bar{Y}$				
	22	**45**				

$$b_2 = \frac{\sum(X_i-\bar{X})(Y_i-\bar{Y})}{\sum(X_i-\bar{X})^2} = \frac{900}{760} = 1.1842$$

$$b_1 = \bar{Y} - b_2\bar{X} = 45 - 1.1842(22) = 18.9476$$

وعليه تكون معادلة الخط المستقيم أو الانحدار البسيط كما يلي:

$$\hat{Y}_i = 18.9476 + 1.1842\ X_i$$

1-2-2 نموذج انحدار بدون حد ثابت

عادة يتضمن الانحدار الحد الثابت، إلا أنه لسبب أو لآخر فقد يقدر أحدهم انحدار بدون حد ثابت، وفي حالة نموذج انحدار بسيط يصبح التوصيف:

$$Y_i = \beta_2 X_i + u_i \tag{2.22}$$

والنموذج المقدّر هو:

$$\hat{Y} = b_2 X_i \tag{2.23}$$

سنشتق صيغة b_2 من المبدأ الأول باستخدام معيار المربعات الصغرى، وتكون البواقي عند المشاهدة i كما يلي:

$$u_i = Y_i - \hat{Y}_i = Y_i - b_2 X_i \tag{2.24}$$

مجموع مربع البواقي هو:

$$SSR = \sum_{i=1}^{n} (Y_i - b_2 X_i)^2 = \sum_{i=1}^{n} Y_i^2 - 2b_2 \sum_{i=1}^{n} X_i Y_i + b_2^2 \sum_{i=1}^{n} X_i^2 \tag{2.25}$$

ونحصل من الشرط الأول على تخفيض $\mathrm{d}SSR/\mathrm{d}b_2 = 0$ كما يلي:

$$\frac{\mathrm{d}SSR}{\mathrm{d}b_2} = 2b_2 \sum_{i=1}^{n} X_i^2 - 2\sum_{i=1}^{n} X_i Y_i = 0 \tag{2.26}$$

وهذا يعطينا:

$$\boldsymbol{b_2 = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2}} \tag{2.27}$$

وتكون المشتقة الثانية $2\sum_{i=1}^{n} X_i^2$ موجبة مؤكدةً على أن SSR أقل ما يمكن، أي أن $0 < \frac{\mathrm{d}^2 SSR}{\mathrm{d}b_2^2} = 2\sum_{i=1}^{n} X_i^2$.

2-3- تطبيق عملي:

طور Okun (1962) العلاقة التطبيقية بين التغيّر في النمو الاقتصادي (التغير في GNP) والتغيّر في معدل البطالة، وسميت هذه العلاقة **بقانون أوكون**، وبيّنت نتائجة أهمية حساسية معدل البطالة لنمو الاقتصاد، وربطت العلاقة الأساسية معدل نمو البطالة (UNEMP) كمتغيّر تابع بحدٍ ثابتٍ ومعدل نمو (GNP) كمتغيّرٍ مستقلٍ كما يلي:

$$\Delta UNEMP_t = a + b\ \Delta GNP_t + u_t$$

ويبيّن الحد الثابت في هذه المعادلة متوسط التغيّر في معدل البطالة عندما يساوي معدل النمو الاقتصادي الصفر؛ ومنها نستنتج أنه عندما لا ينمو الاقتصاد سترتفع البطالة بمعدل a٪، وتشير إشارة المعامل b السالبة إلى أن نمو الاقتصاد يخفّض البطالة، وهي أقل من واحد صحيح؛ فزيادة الناتج القومي الإجمالي GNP بنسبة 1٪ ينخفض معدل البطالة بنسبة b٪. وهذه النتيجة تسمى "**قانون أوكون**".

نأخذ نموذجاً نظرياً لشرح عينة مشاهدات قانون أوكون في الأردن، ونواجه مشكلة استخدام العينة التي تحدد قيم Y_t و X_t لتقدير معلمات الانحدار المجهولة β_1 و β_2؛ وهما معلمة المقطع ومعلمة الميل المجهولتان في علاقة البطالة- النمو الاقتصادي، وإذا استخدمنا 33 فترة زمنية لكل من Y_t و X_t، حيث أن $t = 1, 2, \ldots, N = 33$.

كذلك تواجهنا مشكلة تقدير موقع خط البطالة المتوسط $E(Y) = b_1 + b_2\ X$ الذي يتوسط جميع نقاط البيانات، وهو يعرض متوسط السلوك. ولتقدير b_1 و b_2 سنرسم خط يخترق متوسط البيانات، وبالتالي نستطيع قياس الميل والمقطع بالمسطرة، ومشكلة هذه الطريقة هي اختلاف الخطوط المرسومة بإختلاف الأشخاص وهذا يؤدي إلى رسم خطوط مختلفة، ويؤدي اختلاف المعيار إلى صعوبة تقييم دقة الطريقة.

جدول (2-1) يبيّن الناتج المحلي الاجمالي (القيمة مليون دينار أردني) ومعدّل النمو الاقتصادي الحقيقي ومعدل البطالة خلال الفترة 1982-2014

السنة obs	معدل البطالة U	الناتج المحلي الإجمالي الحقيقي GDPr	معدل النمو الاقتصادي الحقيقي g
1982	4.3	3534.2	7.0
1983	4.8	3455.8	-2.2
1984	5.4	3604.1	4.3
1985	6	3506.5	-2.7
1986	8	3699.5	5.5
1987	8.3	3785.5	2.3
1988	8.8	3840.8	1.5
1989	10.3	3428.7	-10.7
1990	16.8	3419.3	-0.3
1991	15.4	3474.3	1.6
1992	17.5	3972.8	14.3
1993	19.7	4151.1	4.5
1994	18.3	4357.4	5.0
1995	14.6	4627.6	6.2
1996	13.7	4724.2	2.1
1997	12.9	4880.5	3.3
1998	13.7	5027.5	3.0
1999	12.9	5198.0	3.4
2000	13.7	5418.6	4.2
2001	15.8	5704.2	5.3
2002	16.2	6034.1	5.8
2003	15.4	6285.2	4.2
2004	12.4	6823.7	8.6
2005	14.9	7379.6	8.1
2006	14.0	7976.8	8.1
2007	13.1	8629.0	8.2
2008	12.7	9252.1	7.2
2009	12.9	9759.9	5.5
2010	12.5	9985.5	2.3
2011	12.9	10243.8	2.6
2012	12.2	10515.3	2.7
2013	12.6	10812.8	2.8
2014	12.0	11147.6	3.1

المصدر: دائرة الإحصاءات العامة- الأردن.

2-3-1 - مبدأ المربعات الصغرى

هناك عدة صيغ ممكنة لتقدير β_1 و β_2، إلا أننا سنستخدم صيغة تعتمد على **مبدأ المربعات الصغرى Least squares principle** الذي يؤكد على أن الخط المناسب لقيم البيانات يجعل مجموع مربع المسافة الرأسية من تلك النقطة إلى الخط أقل ما يمكن، وهذه المسافة هي مربع القيمة حتى لا يتم حذف المسافة (القيم) الموجبة بالمسافة (القيم) السالبة، وهذه قاعدة احتمالية لكنها فعّالة جداً، وهي إحدى الطرق البسيطة لوصف الخط الذي يخترق متوسط البيانات. ويكون تقدير مقطع وميل هذا الخط أفضل تقدير للبيانات باستخدام مبدأ المربعات الصغرى، وأن b_1 و b_2 هما تقدير المربعات الصغرى للمعلمات β_1 و β_2، وبالتالي يكون الخط المقدّر كما يلي:

$$\hat{Y}_i = b_1 + b_2 X_i \tag{2.28}$$

والمسافة الرأسية من أي نقطة إلى الخط المقدّر هي **بواقي المربعات الصغرى**، وتعطى كما يلي:

$$\hat{u}_i = Y_i - \hat{Y}_i = Y_i - b_1 - b_2 X_i \tag{2.29}$$

2-3-2- تقدير قانون أوكون في الأردن

باستخدام مُقدّرات المربعات الصغرى نستطيع الحصول على تقدير المربعات الصغرى لمعلمات المقطع والميل β_1 و β_2 في مثال قانون أوكون في الأردن باستخدام بيانات الجدول (2-1) نحصل على:

$$b_2 = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2} = -0.872710$$

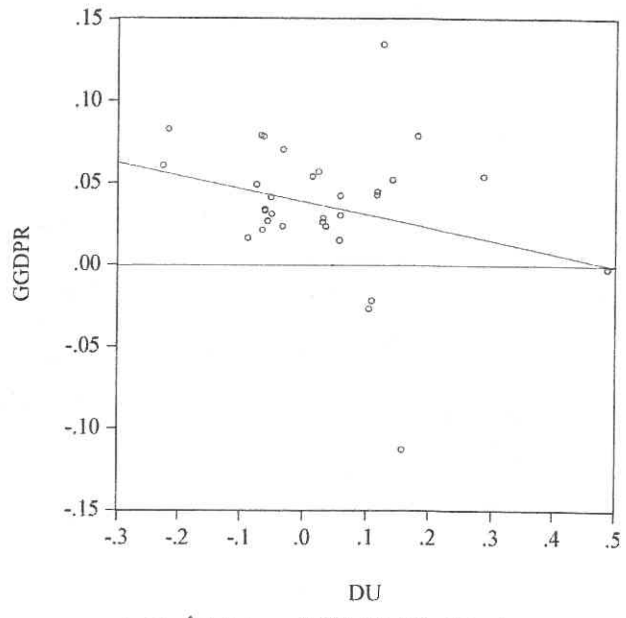
وعلى:

$$b_1 = \overline{Y} - b_2\overline{X} = 0.032072 - (-.872710)(\ 0.035898\ ) = 0.0634$$

والطريقة المناسبة لكتابة قيم b_1 و b_2 لخط الانحدار المقدّر تكون كما يلي:

$$\hat{Y}_t = 0.063 \ - \ 0.873 \ X_t$$

وتم رسم هذا الخط في الشكل (2-5)، وميل الخط هو $0.873 -$ ويقطع مقطعه المحور الرأسي عند 0.06، ويخترق خط المربعات الصغرى المقدّر متوسط البيانات بطريقة دقيقة، وبما أن إحدى خصائص الخط المقدّر تعتمد على تقدير معلمات المربعات الصغرى وتخترق النقاط المعرّفة بوسط العينة $(\overline{X},\ \ \overline{Y}) = (0.035898,\ \ 0.032072)$، ويتبع ذلك صياغة المعادلة كالتالي: $\overline{Y} = b_1 + b_2\ \overline{X}$، وبالتالي تعتبر "نقطة الوسط" قيمة مرجعية مفيدة في تحليل الانحدار.



شكل (2-5) انتشار معادلة أوكون

2-3-3- تفسير معادلة الانحدار

هناك مرحلتين لتفسير معادلة الانحدار: الأولى تحويل المعادلة إلى كلمات يفهمها غير القياسيين، والثاني تقرير ما إذا كان ينبغي أخذ هذا التفسير الحرفي بالقيمة الأسمية، أو ما إذا كان ينبغي مواصلة التحقق من العلاقة، وكلاهما مهم.

Dependent Variable: DU
Method: Least Squares
Sample (adjusted): 1983 2014
Included observations: 32 after adjustments
HAC standard errors & covariance (Bartlett kernel, Newey-West fixed bandwidth = 4.0000)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.063400	0.025770	2.460187	0.0199
GGDPR	-0.872710	0.355098	-2.457662	0.0200

R-squared	0.068936	Mean dependent var	0.032072
Adjusted R-squared	0.037901	S.D. dependent var	0.138071
S.E. of regression	0.135429	Akaike info criterion	-1.100278
Sum squared resid	0.550230	Schwarz criterion	-1.008670
Log likelihood	19.60445	Hannan-Quinn riter.	-1.069912
F-statistic	2.221205	Durbin-Watson stat	2.126941
Prob(F-statistic)	0.146567	Wald F-statistic	6.040102
Prob(Wald F-statistic)	0.019982		

سنشرح قانون أوكون في الأردن خلال الفترة 1983-2014، ونحسب انحدار التغيّر في البطالة dU على نمو الناتج المحلي الإجمالي الحقيقي gGDPr مقيّماً بمليون دينار لـ 33 مشاهدة أخذت من منشورات دائرة الإحصاءات العامة، والتي بينتها النتائج والشكل (2-5) لخط الانحدار أعلاه، وهذه النتيجة تعطينا تقدير لمعامل نمو GDPr وللحد الثابت، وفيما يلي نعرض المعادلة المقدّرة:

$$\Delta\hat{U}_t = 0.063 - 0.873\Delta GDPr_t$$

ويمكن تفسيرها حرفياً كما يلي: يشير معامل الميل إلى أن زيادة نمو GDPr بوحدة واحدة (نسبة مئوية) تنخفض البطالة في الأردن U بمقدار 0.873٪.

صندوق (2-1) تفسير معادلة الانحدار الخطي

نقدم لك طريقة سهلة لتفسير معاملات الانحدار الخطي التالي:

$$\hat{Y}_i = b_1 + b_2 X_i$$

عندما يكون Y و X متغيران بوحدات طبيعية (ليست لوغاريتمية أو على شكل دالة أخرى). في **الخطوة الأولى**: نقول أن زيادة وحدة واحدة من X (مقاسة بوحدات X) تسبب زيادة في Y بمقدار b_2 وحدة (مقاسة بوحدات Y). **والخطوة الثانية**: فحص ماهية وحدات X و Y الفعلية، ونستبدل كلمة "وحدة" بوحدة القياس الفعلي. **والخطوة الثالثة**: إمكانية التعبير عن النتيجة بطريقة أفضل من دون تغيير المضمون.

يبين الثابت b_1 توقع قيمة Y (بوحدات Y) عندما تساوي X الصفر، ومن الممكن أن يكون للحد الثابت معنى معقولاً وقد لا يكون له أي معنى، وهذا يعتمد على السياق الذي ترد فيه.

ماذا عن الحد الثابت؟ أنه يشير إلى المستوى المتوقع من البطالة U عندما يكون نمو $GDPr = 0$، **وفي بعض الأحيان يكون للحد الثابت معنىً واضحاً، وفي بعضها لا يكون له أي معنى**؛ وفي هذه الحالة يكون التفسير الحرفي للحد الثابت بأن متوسط التغيّر في معدل البطالة عندما لا ينمو الاقتصاد (معدل النمو الاقتصادي صفر) يرتفع بمعدل 0.063٪، وقد تكون نتيجته تافهة كما في دالة الاستهلاك في الفصل الأول التي بينت أن الذين لا يملكون أي دينار سيستهلكوا 273.4- مليون دينار، علماً بأن جدول البيانات بيّن أن أقل قيمة استهلاك هي 452.6 مليون دينار. ويقدم الصندوق أدناه دليلاً لتفسير معادلة الانحدار عندما تكون المتغيّرات مقاسة بوحدات طبيعية.

تمرين (1) قدّر معاملات الانحدار الخطي البسيط

يتوفر لديك البيانات التالية:

Y	X	$X_i - \bar{X}$	$Y_i - \bar{Y}$	$(X_i - \bar{X})^2$	$(X_i - \bar{X})(Y_i - \bar{Y})$
2	1				
4	2				
5	3				
4	4				
5	5				

المطلوب: 1- قدّر b_2 و b_1.

2- اكتب نتائجك على شكل معادلة خطية بسيطة (شكل نموذج الانحدار الخطي البسيط).

3- فسر نتائجك.

2-3-4- تغيير وحدات القياس

افرض أن وحدات قياس Y و X تغيّرت. فكيف سيغيّر هذا أثر نتائج الانحدار؟ نبدأ بافتراض النموذج الصحيح التالي:

$$Y_i = \beta_1 + \beta_2 X_i + u_i \tag{2.30}$$

والنموذج المقدّر:

$$\hat{Y}_i = b_1 + b_2 X_i \tag{2.31}$$

نفترض أن وحدات قياس Y قد تغيّرت، وأصبحت بالمقياس الجديد Y* وهو مرتبط بالمقياس القديم:

$$Y_i^* = \lambda_1 + \lambda_2 Y_i \tag{2.32}$$

يشمل تغيّر القياس مقياس المضاعف البسيط مثل تحويل الرطل (الباوند) إلى الغرام، وقد يصادف المرء أحياناً تحويل خطي كامل، تحويل درجات الحرارة من درجة فهرنهايتية إلى درجة مئوية مثلاً، ويكون انحدار Y* على X كما يلي:

$$b_2^* = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i^* - \bar{Y}^*)}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

$$= \frac{\sum_{i=1}^{n}(X_i - \bar{X})([\lambda_1 + \lambda_2 Y_i] - [\lambda_1 + \lambda_2 \bar{Y}])}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

$$= \frac{\sum_{i=1}^{n}(X_i - \bar{X})(\lambda_2 Y_i - \lambda_2 \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

$$= \frac{\lambda_2 \sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \lambda_2 b_2 \qquad (2.33)$$

ينتج عن تغيّر المقياس مُعامل ميلٍ مضروباً بالمقدار λ_2. وأن تغيّر وحدة واحدة في Y يساوي تغيّر λ_2 وحدة في Y*. وحسب معادلة الانحدار، فإن تغيّر X بوحدة يؤدي إلى تغيّر وحدات b_2 في Y، وبالتالي ستؤدي إلى تغيّر Y* بـ $\lambda_2 b_2$ وحدة، أما الأثر على الحد الثابت سيترك كتمرين، وأثر التغيّر في وحدات قياس X سيترك كتمرين.

ومن المهم أن يكون واضحاً في أذهاننا عند تفسير معادلة الانحدار: أن b_1 هي تقدير β_1، و b_2 تقدير β_2، وبذلك يكون التفسير فقط للتقدير. وتتأثر معادلة الانحدار بعامل عشوائي. كما أن التفسير خاص بالمعادلة بعينها.

2-3-5- المرونة ELASTICITY

المرونة هي التغيّر النسبي في أحد المتغيّرات الذي يساهم بتغيّر المتغيّر الآخر، فمرونة Y بالنسبة للتغيّر في X هي:

$$\varepsilon_{YX} = \frac{\Delta Y/Y}{\Delta X/X} = \frac{\Delta Y}{\Delta X} \times \frac{X}{Y} = Slope \times \frac{X}{Y} = b_2 . \frac{X}{Y} \qquad (2.34)$$

فالمرونة حاصل ضرب ميل العلاقة ونسبة القيمة X إلى القيمة Y، أما بالنسبة للعلاقة الخطية وحيث أن الميل ثابت $b_2 = \frac{\Delta Y}{\Delta X}$، فإن المرونة هي التغيّر عند أي نقطة على الخط. وبالتالي، فإن المرونة $\varepsilon_{YX} = b_2\left(\frac{\overline{X}}{\overline{Y}}\right)$، حيث أن التغيّر النسبي في Y هو $\Delta Y/Y$، والتغيّر النسبي في Y هو $\%\Delta Y$، وأن $\%\Delta Y = 100 \times \frac{\Delta Y}{Y}$.

مثلاً تعرّف مرونة الدخل Income elasticity كما يلي:

$$\varepsilon = \frac{\Delta PC/PC}{\Delta GDP/GDP} = \frac{\Delta PC}{\Delta GDP} . \frac{GDP}{PC} = b_2 . \frac{GDP}{PC}$$

والتي تطبق بتعويض القيم المقدّرة محل القيم المجهولة كما يلي:

$$\hat{\varepsilon} = b_2 . \frac{\overline{GDP}}{\overline{CONS}} = 0.75 \times \frac{4696.431}{3247.813} = 1.085$$

2-3-6- التنبؤ PREDICTION

يمكن استخدام المعادلة المقدّرة في عملية التنبؤ prediction أو التوقع forecast، مثلاً لو أردنا التنبؤ بانفاق القطاع العائلي على السلع والخدمات

الاستهلاكية السنوية عندما يكون الدخل السنوي 16087 مليون دينار سنعوض $X = 16087$ في المعادلة المقدّرة ونحصل على:

$$\hat{Y}_i = -273.4 + 0.75\,(16087) = 11791.85$$

سيتم التنبؤ بأن القطاع العائلي عندما يكون دخله السنوي 16087 مليون دينار سينفق منها على استهلاكه 11791.85 مليون دينار.

تمارين

1-2 يظهر الجدول معدل نمو العمل النسبي السنوي e، ومعدل نمو الناتج المحلي الإجمالي الحقيقي g، لخمس وعشرين دولة من دول OECD للفترة 1988-1997، كما وتظهر نتائج انحدار e على g أدناه، **ما هو تفسيرك للمعلمات.**

معدل النمو السنوي لنمو العمالة والناتج المحلي الإجمالي

	e	g		e	g
Australia	1.68	3.04	Korea	2.57	**7.73**
Austria	0.65	2.55	Luxembourg	3.02	**5.64**
Belgium	0.34	2.16	Netherlands	1.88	**2.86**
Canada	1.17	2.03	New Zealand	0.91	**2.01**
Denmark	0.02	2.02	Norway	0.36	**2.98**
Finland	-1.06	1.78	Portugal	0.33	**2.79**
France	0.28	2.08	Spain	0.89	**2.60**
Germany	0.08	2.71	Sweden	-0.94	**1.17**
Greece	0.87	2.08	Switzerland	0.79	**1.15**
Iceland	-1.13	1.54	Turkey	2.02	**4.18**
Ireland	2.12	6.40	UK	0.66	**1.97**
Italy	-0.30	1.68	USA	1.53	**2.46**
Japan	1.06	2.81			

C GDP LS EMPLOYMENT

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.640029	0.295498	-2.165936	0.0409
GDP	0.508389	0.091784	5.538981	0.0000

2-2 إذا كانت $\bar{e} = 0.83$ و $\bar{g} = 2.82$ و $\sum(g_i - \bar{g})^2 = 60.77$ و $\sum(e_i - \bar{e})(g_i - \bar{g}) = 29.76$ ، احسب معاملات الانحدار.

3-2 قدر نتائج انحدار الوزن مقاساً بالرطل على الطول مقاساً بالأنش، واشرح نتائج المعاملات التالية:

Dependent Variable: WEIGHT

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-272855.0	2179.676	-125.1815	0.0000
HEIGHT85	0.066183	0.014894	4.443694	0.0000

4-2 أظهرت نتائج انحدار عدد الأطفال في العائلة على عدد سنوات تعليم الأم أدناه، اشرح المعلمات.

Dependent Variable: Children

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.198478	0.3773667	19.08	0.0000
SM	-0.2525473	0.0316673	-7.98	0.0000

5-2 أثبت أن القيم المقدّرة للمتغيّر التابع غير مرتبطة بالبواقي في نموذج الانحدار البسيط (هذه النتيجة تعميم لحالة الانحدار المتعدد).

2-4- فرضيات نموذج الانحدار الخطي

على افتراض أن نموذج الانحدار الخطي التقليدي الفعلي في العالم الحقيقي هو النموذج التالي: $Y_i = \beta_1 + \beta_2 X_i + u_i$، والنموذج المقدّر له: $\hat{Y}_i = b_1 + b_2 X_i$.

نفترض لهذا النموذج ثمانِ فرضيات أساسية هي كما يلي:

1- الخطية: الفرضية الأولى هي إمكانية حساب المتغيّر التابع كدالة خطية في مجموعة المتغيّرات المستقلة المحددة وفي حد الخطأ، إضافة إلى صحة توصيف المعادلة؛ ونستطيع التعبير عنها رياضياً كما يلي: يكون نموذج الانحدار خطياً في معلمات غير معروفة β_1 و β_2، أي أن: $Y_i = \beta_1 + \beta_2 X_i + u_i$، حيث أن $i = 1, 2, ..., n$. وكذلك خطية معلمات النموذج **Linear in Parameters**؛ أي أن لكل حد على الجانب الأيمن معامل مثل β_i أسه واحد، وعدم وجود علاقة بيّن المعلمات؛ ومثالاً لنموذج معلماته غير خطية يكون $Y = \beta_1 X^{\beta_2} + u$.

2- قيم المتغيّر X_i متغيّرة ويجب أن تأخذ على الأقل قيمتين مختلفتين: تعني هذه الفرضية أن جميع مشاهدات X_t ليست ثابتة وعلى الأقل واحدة منها مختلفة، ويجب أن تكون أكبر من عدد المتغيرات المستقلة، وعليه يكون تباين العينة $var(X)$ غير مساوٍ للصفر $\mathbf{var}(X) \neq 0$، ومن المهمّ التمييز بين تباين العينة الذي يبيّن اختلاف قيم X_i خلالها، والطبيعة العشوائية للمتغيّر X_i. وفي عدة مواقع في هذا الكتاب سنشير إلى أن X_i ليست عشوائية؛ وهذا يعني أن تباين X_i عند أي نقطة يساوي صفر، أي أن $\text{var}(X) = 0$، إلا أنه يوجد قيم مختلفة للمتغيّر X_i خلال العينة. ولمزيد من التوضيح، إذا كانت قيم X ثابتة في العينة، فإنه لا يمكن حساب الانحرافات عن المتغيّر Y، وإذا حاولنا تقدير انحدار Y على X عندما تكون X ثابته، فإننا لا نستطيع حساب معاملات الانحدار؛ لأن

المتغيّر X سيساوي متوسطه $\bar{X}$ لجميع i، وبالتالي سيساوي بسط ومقام المعلمة b_2 الصفر، وبما أننا لا نستطيع حساب b_2 فإننا لا نستطيع الحصول على b_1.

3- المتغيّر X_t غير عشوائي non-stochastic ويظهر في العينات المكررة: تعني هذه الفرضية أن X_t متغيّر وقيمه ليست محددة بأي آلية مصادفة chance، وهو محدد من قِبل المُجرِّب أو الباحث. ثمّ من الممكن الحصول على نفس قيم المتغيّر المستقل عند تكرار العينة، وهذا يعني أن $\mathrm{cov}(X_i,u_j)=0$ لجميع قيم i و $j = 1, 2, ..., n$، وأن X_i و u_j غير مرتبطين.

4- القيمة المتوقعة لحد الخطأ تساوي صفر: نفترض أن القيم المتوقعة لحد الخطأ لأي مشاهدة ينبغي أن تساوي الصفر، ويكون حد الخطأ في بعض الأحيان موجباً وفي بعضها سالباً، لكن لا ينبغي أن يكون لها اتجاهاً منتظماً في تحركها؛ وهذا يعني أن حد الخطأ حقيقي، فإذا أخذنا عينة كبيرة الحجم سيساوي الوسط الحسابي لحدود الخطأ صفراً، وهذا يشير إلى أن $E(u)=0$، ونحتاج لهذه الفرضية لتفسير الجزء المحدّد من نموذج الانحدار $b_1+b_2X_i$.

5- تجانس تباين الخطأ Homosekedasticity: نفترض أن تباين حد الخطأ ثابت؛ وهذا يتطلب أن يكون لحد الخطأ نفس التباين (أي أن: ثابت $\mathrm{var}(u_i)=\sigma^2=$ لجميع i). وبما أن σ_u^2 مجهولة، يتطلب تحليل الانحدار تقدير تباين حد الخطأ. وإذا كانت هذه الفرضية غير مقبولة عندها تكون معاملات انحدار OLS غير كفؤة، ومن الممكن الحصول على نتائج انحدار أكثر ثقة بتعديل طريقة الانحدار.

6- استقلال قيم حد الخطأ Serial Independence: هذا يتطلب أن يكون توزيع جميع حدود الخطأ مستقلاً، أو غير مرتبط أحدها بالأخرى (u_i

مستقلة عن u_j، حيث أن ($j \neq i$)، وعليه يكون التباين المشترك لأي زوج من الأخطاء العشوائية يساوي الصفر، أي أن: $\mathbf{cov}(u_i, u_j) = 0$ لجميع قيم $i \neq j$، ويعني هذا الشرط أن حدود الخطأ في أي فترة يجب أن لا تكون مرتبطةً بحد الخطأ في الفترة اللاحقة أو السابقة. وبعبارة أخرى، نفترض أن حد الخطأ يخلو من الارتباط الذاتي **autocorrelation**؛ وهذا يعني خلو أي مشاهدتين من ارتباط منتظم بيّنها، ويجب أن تكون قيم حد الخطأ مستقلة عن الأخرى، وإذا كانت هذه الفرضية غير مقبولة يكون تقدير OLS غير كفؤ.

إذا تحققت الفرضيات (1-6) تكون المقدّرات β_1 و β_2 المقدرة بطريقة المربعات الصغرى العادية *OLS* "افضل مُقدّرات خطية غير متحيّزة": Best (most efficient) Linear (function of observation on Y) Unbiased Estimators (BLUE) ويعطينا مجموع مربع البواقي مقسوماً على عدد درجات الحرية مقدَّراً غير منحاز لتقدير التباين σ_u. وماذا يعني هذا التعبير BLUE؟

- "**المقدّرات** Esimator" $\hat{\beta}_1$ و $\hat{\beta}_2$.هي قيم صحيحة للمعلمات β_1 و β_2.
- "**خطّي** Linear" تعني مقدرات $\hat{\beta}_1$ و $\hat{\beta}_2$ خطية أو صيغة $\hat{\beta}_1$ و $\hat{\beta}_2$ مزيج خطي لمتغيّرات عشوائية (Y).
- "**غير متحيّز** Unbiased" تساوي قيم $\hat{\beta}_1$ و $\hat{\beta}_2$ بالمتوسط قيمها الفعلية الصحيحة.
- "**أفضل** Best" تعني أن مقدرات OLS للمعلمة $\hat{\beta}_2$ لها أقل تباين بين المقدرات الخطية غير المتحيّزة.

7- توزيع البواقي طبيعي *Normality of residuals*: يفترض أن تكون حدود الخطأ $u_1, u_2, \cdots, u_n$ مستقلة وتوزيعها طبيعي متماثل، بوسط صفر وتباين ثابت. وكذلك يكون توزيع معاملات الانحدار طبيعياً إذا كان حد الخطأ توزيعه طبيعي في كل مشاهدة. فإذا كان توزيع u طبيعياً ستكون نتائج الانحدار مفيدة لتطبيق اختبار t واختبار F للفرضيات وتكوين فترات ثقة β_1 و β_2.

8- عدم وجود ارتباط خطي متعدد ***Multicollinearity***: عدم وجود ارتباط خطي متعدد تام بين المتغيّرات المستقلة.

2-4-1- انتهاك الفرضيات

توضح أول ثلاث فرضيات أن X_i متغيّر سلوكه لم يكن مختاراً بالصدفه، ونستطيع اختياره بالتكرار، لأن X_i يستخدم لشرح ما يحدث (متغيّر تفسيري).

يخلق انتهاك الفرضية 1 مشكلة تسمى خطأ التوصيف مثل متغيّرات تفسيرية خاطئة وعدم الخطية، وانتهاك الفرضية 2 و 3 يظهر في أخطاء المتغيّرات، ويقود انتهاك الفرضية 4 إلى انحياز المقطع (الحد الثابت)، بينما انتهاك الفرضية 5 و 6 يؤدي إلى مشاكل اختلاف التباين *Heteroskedasticity* والارتباط المتسلسل *serial correlation* على التوالي، والفرضية 7 لها آثار مهمة في اختبار الفرضية، وانتهاك الفرضية 8 يؤدي إلى مشكلة الارتباط الخطي المتعدد التام *Multicollinearity*.

جدول (2-2) فرضيات نموذج الانحدار الخطي التقليدي

الفرضية	الصيغة الرياضية	ماذا يعني انتهاك الفرضية	الفصل
1- خطية النموذج	$Y_i = \beta_1 + \beta_2 X_i + u_i$	متغيّرات تفسيرية خاطئة عدم الخطية تغيّر المعلمات	3 4 3
2- X متغيّر	$\text{var}(X) \neq 0$	أخطاء المتغيّرات	3
3- X غير عشوائي، وثابت في تكراره	$\text{cov}(X_i, u_j) = 0$	ارتباط ذاتي Autocorrelation	7
4- القيمة المتوقعة للتوزيع تساوي صفر	$E(u) = 0$	تحيّز الحد الثابت (المقطع)	-
5- تجانس التباين	$\text{var}(u_i) = \sigma^2 =$ ثابت	اختلاف التباين	6
6- الاستقلال المتسلسل	$\text{cov}(u_i, u_j) = 0 \quad , i \neq j$	ارتباط ذاتي	7
7- التوزيع الطبيعي	$u_i \sim N(\mu, \sigma^2)$	قيم متطرفة	3
8- علاقات ارتباط غير خطية		ارتباط خطي متعدد	5

2-5- خصائص مقدّرات المربعات الصغرى العادية

اعتماداً على فرضيات نموذج الانحدار الخطي التقليدي نستطيع اثبات أن مقدرات المربعات الصغرى العادية هي أفضل تقدير خطي غير متحيز *Best Linear Unbiased Estimators (BLUE)*؛ ولإجراء ذلك يجب علينا أولاً تحليل معاملات الانحدار المقدّرة بطريقة المربعات الصغرى العادية *(OLS)* إلى مكونات غير عشوائية وعشوائية.

كنقطة بداية نلاحظ أن المعادلة Y_i لها (1) مكوّن غير عشوائي $\alpha + \beta X_i$ يعتمد على X وعلى المعلمات، (2) وعلى مكوّن عشوائي تلتقطه البواقي u_i؛ حيث تعتمد β بشكل غير مباشر على u_i. علماً بأن تقدير المعادلة هو $\hat{Y} = b_1 + b_2 X$. فإذا كانت قيم u_i في عينة مختلفة، سيكون لدينا قيم مختلفة من Y، وبالتالي قيم مختلفة من b_2، ويمكننا نظرياً تحليل b_2 إلى مكونات غير عشوائية وعشوائية، وتكون الخطوة الأولى التعويض في Y ووسط العينة من النموذج الحقيقي. ويتم حذف β_1 ونعيد ترتيب الحدود المتبقية.

$$b_2 = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2}$$

$$= \frac{\sum(X_i - \bar{X})([\beta_1 + \beta_2 X_i + u_i] - [\beta_1 + \beta_2 \bar{X} + \bar{u}])}{\sum(X_i - \bar{X})^2}$$

$$= \frac{\sum(X_i - \bar{X})(\beta_2(X_i - \bar{X}) + u_i - \bar{u})}{\sum(X_i - \bar{X})^2}$$

$$= \frac{\sum \beta_2 \left(X_i - \overline{X}\right)^2 + \left(X_i - \overline{X}\right)\left(u_i - \overline{u}\right)}{\sum \left(X_i - \overline{X}\right)^2}$$

$$= \beta_2 + \frac{\sum \left(X_i - \overline{X}\right)\left(u_i - \overline{u}\right)}{\sum \left(X_i - \overline{X}\right)^2} \qquad (2.35)$$

وبالتالي يكون معامل b_2 المقدّر بطريقة المربعات الصغرى من أي عينة له مكوّن غير عشوائي β_2 ومكوّن عشوائي يعتمد على $\sum \left(X_i - \overline{X}\right)\left(u_i - \overline{u}\right)$.

أ- الخطية Linearity

بالاعتماد على الفرضة (3) يكون X غير عشوائي وثابت في العينات المكررة، ونستطيع معالجة قيم X كثوابت في العينات، ونحتاج فقط التركيز على قيم Y، فإذا كانت مقدّرات المربعات الصغرى العادية دالة خطية في قيم Y تكون مقدرات خطية، ومن المعادلة التالية:

$$b_2 = \frac{\sum \left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sum \left(X_i - \overline{X}\right)^2}$$

$$= \frac{\sum \left(X_i - \overline{X}\right)Y_i - \overline{Y}\left(X_i - \overline{X}\right)}{\sum \left(X_i - \overline{X}\right)^2} \qquad (2.36)$$

وبما أن مجموع $\overline{Y}\left(X_i - \overline{X}\right) = 0$ فإن:

$$b_2 = \frac{\sum(X_i - \overline{X})Y_i}{\sum(X_i - \overline{X})^2} = \sum z_i Y_i \qquad (2.37)$$

حيث أن $z_i = \sum[(X_i - \overline{X})/(X_i - \overline{X})^2]$ نستطيع اعتباره ثابتاً، وبالتالي تكون b_2 مقدَّراً خطياً للمتغيّر Y_i.

ب- الاتساق Consistency

تكون مقدرات المربعات الصغرى b_1 و b_2 متسقة عندما تكون احتمالية الفرق بين القيمة المقدّرة والقيمة الصحيحة في النهاية صفراً، وبالتالي يقترب التقدير من قيمته الصحيحة عندما يزداد حجم العينة إلى ما لا نهاية؛ وبعبارة أخرى يعني الاتساق ببساطة أن تقدير b سيساوي β الصحيحة؛ وهذا يعني أن التقدير يذهب إلى ما لا نهاية وستتقارب b من القيمة الصحيحة $\text{Plim}(b) = \beta$ وتكون الفرضية $E(X_t u_t) = 0$ و $E(u_t) = 0$ كافية لاشتقاق اتساق مقدرات OLS.

ج- عدم التحيّز Unbiasedness

يكون تقدير b_2 مقدَّراً غير متحيّز للمعلمة β_2 عندما تكون:

$$E(b_1) = \beta_1$$

و

$$E(b_2) = \beta_2$$

بالمتوسط ستساوي القيم المقدرة للمعاملات قيمها الصحيحة، ولا تكون أكبر ولا أقل من تقدير المعاملات الصحيحة، وهذا يتطلب $\text{cov}(X_t, u_t) = 0$، وبالتالي سيكون تقدير المربعات الصغرى تقديراً غير متحيّز على فرض أن المتغيّر التفسيري X_t مستقلاً عن حد الخطأ u_t. ويتضح أن عدم التحيّز شرط أقوى من الاتساق، حيث يتحقق في جميع العينات الصغيرة منها والكبيرة، فالمقدّر المتسق قد يكون متحيّزاً في العينات الصغيرة، فهل جميع المتغيّرات غير المتحيّزة هي متسقة؟ في الحقيقة لا، المقدرات غير المتحيّزة تكون متسقة إذا انخفض تباينها عندما يزداد حجم العينة.

د- الكفاءة وأفضل تقدير خطي غير متحيّز

نستطيع من الفرضية 5 و 6 اثبات أن مقدَّرات المربعات الصغرى العادية أكثر كفاءة بين جميع المقدَّرات الخطية غير المتحيّزة، ونستنتج أن إجراء المربعات الصغرى العادية ينتج مقدَّرات BLU.

إن اثبات مقدَّرات المربعات الصغرى العادية هي مقدَّرات BLU معقّدة نسبياً، حيث نبدأ بالتقدير ونحاول اشتقاق مقدرات BLU للمعلمة β_2 بالاعتماد على الخصائص الخطية وعدم التحيّز وأدنى تباين، ثم نفحص فيما إذا كان مقدر BLU مشتق بهذا الإجراء هو نفس مقدر المربعات الصغرى العادية.

2-6- نظرية غاوس-ماركوف THE GAUSS-MARKOV THEOREM

ماذا نستطيع أن نقول عن مُقدَّرات المربعات الصغرى b_1 و b_2؟

- تستخدم الصيغة (2.14) و (2.21) لتقدير المعلمات المجهولة β_1 و β_2 في نموذج الانحدار الخطي البسيط.

- مُقدَّرات المربعات الصغرى هي مُقدَّرات خطية، ويمكن كتابة كل من b_1 و b_2 كمتوسط مرجح لقيم Y_1.
- توفر الفرضيات 1-6 مُقدَّرات المربعات الصغرى غير المتحيزة؛ وهذا يعني أن $E(b_1)=\beta_1$ و $E(b_2)=\beta_2$.
- لدينا صيغة تباين وتباين مشترك للمعلمات b_1 و b_2 ولدينا نقاش عن عدم تحيّز المقدَّرات، ووجود أصغر تباين هو الأفضل؛ وهذا يعني أنه لدينا فرصة كبيرة للحصول على تقدير قريب جداً من القيم الصحيحة للمعلمات.

الآن سنناقش نظرية غاوس- ماركوف الشهيرة، في ظل الفرضيات 1-6 لنموذج الانحدار الخطي، ويكون للمُقدَّرات b_1 و b_2 أصغر تباين لجميع المقدّرات β_1 و β_2 الخطية وغير المتحيّزة، وتكون β_1 و β_2 مُقدَّرات خطية غير متحيّزة ***best linear unbiased estimators (BLUE)***.

لنرى ماذا تقول نظرية غاوس ماركوف:

1- تكون المقدّرات b_1 و b_2 "الأفضل" عند مقارنتها بمُقدَّرات مماثلة، وتكون خطية وغير متحيّزة، ولا تقول النظرية أن b_1 و b_2 هي الأفضل لجميع المقدّرات الممكنة.

2- المقدّرات b_1 و b_2 هي الأفضل لأن لها تباين أقل، وعندما نقارن مُقدَّرات خطية وغير متحيّزة، نريد دائماً أحدها بأقل تباين، حيث أن صيغة التقدير تعطي أعلى احتمالية للحصول على تقدير قريب من قيمة المعلمة الصحيحة.

3- للحفاظ على نظرية غاوس-ماركوف يجب أن تكون الفرضيات 1-6 صحيحة، وإذا كان أحدها غير صحيح، عندها لا تكون b_1 و b_2 أفضل مُقدَّرات خطية غير متحيّزة للمعلمات β_1 و β_2.

4- لا تعتمد نظرية غاوس-ماركوف على فرضية الطبيعية (الفرضية 7).

5- تُطبّق نظرية غاوس-ماركوف على مُقدّرات المربعات الصغرى، ولا تُطبّق على تقدير المربعات الصغرى لعينة فردية.

2-7- التوزيع الاحتمالي لمُقدّرات المربعات الصغرى

طُوِّرت خصائص مُقدّرات المربعات الصغرى على أساس عدم اعتمادها على فرضية الطبيعية، أما إذا افترضنا أن الأخطاء العشوائية (u_i) توزيعها طبيعي بوسط حسابي صفر وتباين ثابت σ^2، سيكون التوزيع الاحتمالي طبيعي لمُقدّرات المربعات الصغرى كذلك. وحصلنا على هذه النتيجة بخطوتين: الأولى تعتمد عل فرضية 1 $\left[E(Y|X)=b_1+b_2X\right]$، فإذا كانت u_i طبيعية تكون Y_i كذلك، والثانية أن مُقدّرات المربعات الصغرى خطية $b_2=\sum w_iY_i$، وتوزيع مجموع المتغيّرات العشوائية طبيعي، وإذا افترضنا الطبيعية (الفرضية 7 لحد الخطأ) سيكون توزيع مُقدّرات المربعات الصغرى طبيعياً:

$$b_1 \sim N\left(\beta_1, \frac{\sigma^2\sum X_i^2}{N\sum\left(X_i-\bar{X}\right)^2}\right) \tag{2.38}$$

$$b_2 \sim N\left(\beta_2, \frac{\sigma^2}{\sum\left(X_i-\bar{X}\right)^2}\right) \tag{2.39}$$

إن طبيعية مُقدّرات المربعات الصغرى مهمة جداً، لكن ماذا لو لم يكن توزيع الأخطاء طبيعياً؟

صندوق (2-2) نظرية A Central Limit Theorem

إذا تحققت الفرضيات (1-6) وكان حجم العينة N كبيراً بما فيه الكفاية يكون توزيع مُقدّرات المربعات الصغرى طبيعياً تقريباً.

ما هو الحجم المناسب للعينة ليكون كبيراً بما فيه الكفاية؟ الجواب لا يوجد رقم محدد، والسبب أن هذا الجواب غير دقيق وغير مُرْضٍ أن "الحجم" يعتمد على عدة عوامل مثل أن يكون توزيع الأخطاء عشوائياً، وهل هو منتظم، أم ملتوٍ، وماذا تشبه قيم X_i. نقول في نموذج الانحدار البسيط أن $N = 30$ عددٌ كافٍ، وفي بعضها نقول أن $N = 50$ سيكون كافياً، وأن معنى "العدد كافٍ" يتغيّر من مسألةٍ إلى مسألة.

2-8- تقدير تباين حد الخطأ

تباين حد الخطأ العشوائي σ^2 في نموذج الانحدار الخطي البسيط هو معامل مجهول وعلينا تقديره كما يلي:

$$\mathrm{var}(u_i) = \sigma^2 = \mathrm{E}[u_i - \mathrm{E}(u_i)]^2 = \mathrm{E}(u_i^2) \quad (2.40)$$

فإذا كانت الفرضية $\mathrm{E}(u_i) = 0$ صحيحة، أي أن "التوقع" هو متوسط القيمة التي نقدرها (σ^2) كمتوسط مربع الأخطاء:

$$\hat{\sigma}^2 = \frac{\sum u_i^2}{N} \quad (2.41)$$

وبما أن الأخطاء العشوائية مجهولة علينا الحصول على ما يناظرها، والمسماة ببواقي المربعات الصغرى، وهي:

$$u_i = Y_i - b_1 - b_2\ X_i \qquad (2.42)$$

ويتم الحصول على بواقي المربعات الصغرى بتعويض المعلمات المجهولة بتقدير المربعات الصغرى لها.

$$\hat{u}_i = Y_i - \hat{Y}_i = Y_i - b_1 - b_2\ X_i \qquad (2.43)$$

وبالتالي فإن:

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{N} \qquad (2.44)$$

وهذا التقدير في العينة الكبيرة هو مُقدّر متحيّز للتباين σ^2، وبإجراء تعديل بسيط ينتج مُقدَّر غير متحيّز كما يلي:

$$\boldsymbol{\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{N-2}} \qquad \textbf{(2.45)}$$

يتم طرح 2 من المقام؛ وهو عدد معلمات الانحدار (b_1 و b_2) في النموذج يجعل تقدير $\hat{\sigma}^2$ غير متحيّز، وبالتالي فإن $E(\hat{\sigma}^2) = \sigma^2$.

مثال

أحسب تباين حد الخطأ العشوائي في مثال نموذج الانحدار الخطي البسيط السابق، بالاعتماد على الجدول أدناه.

الحل

X	Y	$\hat{Y}_i$	$Y_i - \hat{Y}_i$	$(Y_i - \hat{Y}_i)^2$
20	30	42.6316	-12.6316	159.5573
40	60	66.3156	-6.3156	39.8868
20	40	42.6316	-2.6316	6.925319
30	60	54.4736	5.5264	30.5411
10	30	30.7896	-0.7896	0.623468
10	40	30.7896	9.2104	84.83147
20	40	42.6316	-2.6316	6.925319
20	50	42.6316	7.3684	54.29332
20	30	42.6316	-12.6316	159.5573
30	70	54.4736	15.5264	241.0691
			المجموع	**784.2105**

وعليه يكون تباين حد الخطأ يساوي:

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{N-2} = \frac{\sum (Y_i - \hat{Y})^2}{N-2} = \frac{784.2105}{10-2} = \frac{784.2105}{8} = 98.0263$$

2-8-1- تقدير التباين والتباين المشترك لمُقدّرات المربعات الصغرى

إذا كان لديك مُقدّرات غير متحيّزة لتباين الخطأ، فإن هذا يعني أننا نستطيع تقدير تباين مُقدّرات المربعات الصغرى b_1 و b_2 والتباين المشترك بينهما.

$$\hat{\text{var}}(b_1) = \hat{\sigma}^2 \left[\frac{\sum X_i^2}{N \sum (X_i - \bar{X})^2} \right] \quad (2.46)$$

$$\hat{\text{var}}(b_2) = \frac{\hat{\sigma}^2}{\sum (X_i - \bar{X})^2} \qquad (2.47)$$

$$\hat{\text{cov}}(b_1, b_2) = \hat{\sigma}^2 \left[\frac{-\bar{X}}{\sum (X_i - \bar{X})^2} \right] \qquad (2.48)$$

وعند أخذ الجذر التربيعي للتباين المقدّر نحصل على الخطأ المعياري **standard errors** للمعلمة b_1 و b_2، وتستخدم هاتين القيمتين لاختبار الفرضيات وفترات الثقة، ويشار إليهما $se(b_1)$ و $se(b_2)$:

$$se(b_1) = \sqrt{\hat{\text{var}}(b_1)} \qquad (2.49)$$

$$se(b_2) = \sqrt{\hat{\text{var}}(b_2)} \qquad (2.50)$$

مثال

قدر تباين مُقدّرات المربعات الصغرى b_1 و b_2 والتباين المشترك بينهما للمثال السابق، علماً بأن التباين لحد الخطأ العشوائي يساوي 98.0263.

الحل

نتبع الخطوات التالية:

X_i	Y_i	$X_i - \bar{X}$	$(X_i - \bar{X})^2$	X^2
20	30	-2	4	400
40	60	18	324	1600
20	40	-2	4	400
30	60	8	64	900
10	30	-12	144	100
10	40	-12	144	100
20	40	-2	4	400
20	50	-2	4	400
20	30	-2	4	400
30	70	8	64	900
		$\sum$	**760**	**5600**
$\bar{X} = 22$				

باستخدام بيانات الجدول أدناه نستطيع حساب ما يلي:

$$\hat{\text{var}}(b_1) = 98.0263\left[\frac{5600}{10(760)}\right] = 92.2299$$

$$se(b_1) = \sqrt{\hat{\text{var}}(b_1)} = \sqrt{92.2299} = 9.6036$$

$$\hat{\text{var}}(b_2) = \frac{\hat{\sigma}^2}{\sum (X_i - \bar{X})^2} = \frac{98.0263}{760} = 0.1289$$

$$se(b_2) = \sqrt{\hat{\text{var}}(b_2)} = \sqrt{0.1289} = 0.3590$$

$$\hat{\text{cov}}(b_1, b_2) = \hat{\sigma}^2 \left[\frac{-\bar{X}}{\sum (X_i - \bar{X})^2}\right] = 98.0263 \times \left[\frac{-22}{760}\right] = -2.8376$$

2-9- اختبار الفرضيات HYPOTHESIS TESTS لمعاملات الانحدار

الكثير من مشاكل الأعمال والاقتصاد تتطلب قراراً حول قيم المعلمات؛ ففي مثالنا عن الطلب على السلع أو دالة الاستهلاك يوجد عدة قرارات متعلقة بالمعلمة b_2 التي تشير إلى زيادة الإنفاق على السلع الاستهلاكية بمقدار 75 دينار عندما يزيد الدخل بمقدار 100 دينار، كذلك الاعتقاد بأن β_2 موجبة اعتماداً على النظرية الاقتصادية، ويتم اختبار البيانات والنموذج لبيان فيما إذا كانتا تدعمان هذه النظرية.

وبذلك تكون الخطوة التالية الخروج بفرضية؛ وهي ببساطة تخمين قابل للاختبار والإجابة على سؤال البحث، وغالباً ما توصف الفرضية بأنها محاولة الباحث لتفسير الظاهرة المثيرة للاهتمام، ويمكن أن تتخذ الفرضيات أشكالاً مختلفة، اعتماداً على أن السؤال الذي يطرح ونوع الدراسة التي تجريها. وتأخذ الفرضيات صيغة "إذا- فإن، *if-then*" في أبسط أشكالها؛ فعلى سبيل المثال قد يفترض الباحث أنه "إذا زاد الدخل بمقدار 100 دينار، فإن الاستهلاك سيزيد بأقل من ذلك".

وقبل أن نناقش أنواع محددة من الفرضيات، هناك نوعان من النقاط المهمّة التي يجب أن نأخذها في الاعتبار: أولاً يجب أن تكون كل الفرضيات قابلة للخطأ، وبالتالي يجب أن تكون الفرضيات قابلة للدحض على أساس نتائج الدراسة. بكل بساطة، إذا كانت فرضية الباحث لا يمكن دحضها، فإن الباحث لا يستطيع إجراء تحقيقه العلمي. والثانية يجب أن تستطيع

الفرضية التنبؤ (عادة عن العلاقة بين متغيّرين اثنين أو أكثر). ويمكن بعد ذلك إما اعتماد الفرضية أو دحضها.

وبالتالي تكون إجراءات اختبار الفرضية بمقارنة التخمين حول المجتمع بالمعلومات الواردة في بيانات العينة، وتُشَكّل الفرضية **Hypotheses** السلوك الاقتصادي في النموذج الاقتصادي والاحصائي، وتبيّن هذه الافتراضات حالة معاملات النموذج، وتستخدم معلومات المعاملات والانحراف المعياري لصياغة الاستنتاج حول الفرضية. وعند اختبار الفرضية يجب استعراض العناصر التالية:

صندوق (2-3) مكونات اختبار الفرضية

بمعرفة توزيع المعاملات المقدّرة نستطيع إجراء اختبار الفرضية لتقييم معنويتها الإحصائية باتباع الخطوات التالية:

1- **تحديد الفرضية الأساسية** H_0 **والبديلة** H_1: إما أن تكون $H_0: \beta_2 = 0$، $H_1: \beta_2 \neq 0$ (اختبار بذيلين)، أو تكون عندما يكون لدينا معرفة مسبقة عن إشارة المعامل المقدّر (مثل الإشارة الموجبة) $H_0: \beta_2 = 0$، $H_1: \beta_2 > 0$ (اختبار بذيل واحد).

2- **احسب إحصائية** t حسب الصيغة $t = \dfrac{b_2 - \beta_2^{\circ}}{s.e.(b_2)}$؛ وحيث أن $\beta_2^0 = 0$

تعتمد على الفرضية الأساسية تصبح الصيغة $t = \frac{b_2}{s.e.(b_2)}$.

3- **حدّد قيمة t الحرجة من جدول t** (جدول 1 في الملحق) بدرجات حرية N-2.

4- **الاستنتاج:** إذا كانت $|t_{\text{الحرجة}}| < |t_{\text{المحسوبة}}|$ نرفض الفرضية الأساسية.

ملاحظة: إذا أردنا اختبار فرضية مختلفة كأن تكون $\beta_2^0 = 1$، سنحتاج تغيير فرضيتنا الأساسية والبديلة في الخطوة (1) ونحسب إحصائية t باستخدام صيغة $t = \frac{b_2 - \beta_2^\circ}{s.e.(b_2)}$.

في حالة العينات الكبيرة (تكون درجات الحرية أكبر من 30) نستطيع عدم الرجوع إلى الجداول الإحصائية، وستكون القيمة الحرجة عند مستوى معنوية 5٪، ولعينات كبيرة جداً $N \to \infty$ تصل قيمة إحصائية t إلى ± 1.96. وعند نفس مستوى المعنوية و 30 درجة حرية تكون $t_{\text{الحرجة}} = \pm 2.045$، بينما عند 60 درجة حرية تكون ± 2.00 بالضبط، وبالتالي لعينات كبيرة يكون استخدام القيمة الحرجة $|t| > 2$ آمناً تماماً. ولعينات صغيرة يجب استخدام القيمة المحددة في جدول t حسب القاعدة أعلاه.

يرتبط أداء اختبار الفرضيات بمعاملات الانحدار، وعلى فرض أن النموذج الصحيح هو $Y_i = \beta_1 + \beta_2 X_i + u_i$ وقدّرنا معادلة الانحدار $\hat{Y}_i = b_1 + b_2 X_i$، سنفترض وجود فرضية تتعلق بمعاملات الميل $H_0 : \beta_2 = \beta_2^0$ (β_2^0 هي القيمة المفترضة وغالباً تكون صفراً (0) وهذا يعتمد على النظرية وطبيعة الفرضية)، وسنرفض H_0 إذا كان الفرق بيّن b_2 و β_2° كبيراً جداً مقاساً بالخطأ المعياري، وعندها نعرّف احصائية t كما يلي:

$$t = \frac{b_2 - \beta_2^{\circ}}{s.e.(b_2)} \qquad (2.51)$$

سيتم رفض الفرضية الأساسية H_0 إذا كانت $|t| > t_{\text{الحرجة}}$، أي أن $t > t_{\text{الحرجة}}$ أو $t < -t_{\text{الحرجة}}$.

وعند تشكيل اختبار t بافتراض أن $H_0 : \beta = \beta^0$ ودرجات الحرية تساوي $n-1$، و n هي عدد المشاهدات في العينة، وفي معادلة الانحدار يستهلك تقدير كل معلمة درجة حرية واحدة من العينة، وبالتالي فإن عدد درجات الحرية سيساوي عدد المشاهدات في العينة ناقصاً عدد المعلمات المقدَّرة؛ والمعلمات هي: الحد الثابت ومعاملات المتغيّرات التفسيرية، وفي هذه الحالة، يتضمن تحليل الانحدار البسيط تقدير معلمتين b_1 و b_2، وبالتالي سيكون عدد درجات الحرية $n-2$. وتم التركيز على الصيغة العامة؛ لأن هذا ما يتطلبه تحليل الانحدار المتعدد لاحقاً.

افرض أن نسبة معدل تضخم الاسعار في الاقتصاد (P) تعتمد على نسبة تضخم معدل الأجور (W) حسب المعادلة الخطية التالية:

$$P=\beta_1+\beta_2\, w+u \tag{2.52}$$

حيث أن β_1 و β_2 معاملات المعادلة، و u حد الخطأ، ونفترض (بغض النظر عن تأثير حد الخطأ) أن معدل تضخم الأسعار يساوي معدل تضخم الأجور؛ أي أن زيادة الأجور تزيد الكلفة بشكل متناسب مع زيادة الأسعار، وبالتالي تكون الفرضية الأساسية $H_0:\beta_2=1$، وتكون الفرضية البديلة $H_1:\beta_2\neq 1$، أو على فرض أننا نأخذ مشاهدات فعلية عن متوسط معدّل تضخم الأسعار وتضخم الأجور خلال آخر خمس سنوات لعينة 20 دولة وقدرنا النموذج التالي:

$$\hat{P}=\underset{(0.05)}{-1.21}+\underset{(0.10)}{0.82}\, w \tag{2.53}$$

حيث تمثل الأرقام بين الأقواس الخطأ المعياري، وبالتالي تكون قيمة إحصائية t لاختبار الفرضية الأساسية $H_0:\beta_2=1$ كما يلي:

$$t=\frac{b_2-\beta_2^{\circ}}{s.e.(b_2)}=\frac{0.82-1.00}{0.10}=-1.80 \tag{2.54}$$

وحيث يوجد لدينا 20 مشاهدة في العينة؛ سيكون عدد درجات الحرية (18) درجة وتكون قيمة t الحرجة عند مستوى معنوية 5٪ تساوي 2.101، وبالتالي تكون القيمة المطلقة لإحصائية t أقل من القيمة الحرجة، وفي هذه الحالة لا نستطيع رفض الفرضية الأساسية، كما أن التقدير 0.82 هو أقل من القيمة المفترضة (1)؛ إلا أنها ليست بعيدة كثيراً لكي نستبعد امكانية صحة الفرضية الأساسية.

اختبرنا في هذا المثال فرضية محددة هي أن $H_0:\beta_2=1$ افترضتها النظرية الكينزية، ومن الناحية التطبيقة يكون النموذج النظري التالي $Y_i=\beta_1+\beta_2\, X_i+u_i$.

ومن النادر إفتراض أن المتغيّر X له تأثير محدد على المتغيّر الآخر Y. ويصبح الهدف من اختبار t أقل طموحاً، إلا أن الهدف هو تحديد فيما إذا كان المتغيّر X له تأثير على Y. ونحن نعتقد أن X له تأثير على Y وأن β_2 ليست صفراً، إلا أننا لا نستطيع توقع قيمة المعلمة. لذلك سنعتمد استراتيجية عكسية، ونجعل الفرضية الأساسية $H_0 : \beta_2 = 0$. ونأمل بإثبات وجوب رفض H_0. فإذا نجحنا برفضها سنهتم بحجم الأثر وحد الخطأ للتقدير.

خذ على سبيل المثال دالة الاستهلاك البسيطة:

$$CONS = \beta_1 + \beta_2\ GDP + u \qquad (2.55)$$

حيث أن $CONS$ الاستهلاك الخاص، و GDP الناتج المحلي الإجمالي. وتتوقع النظرية أن زيادة الدخل ستؤدي على زيادة الاستهلاك؛ لكن بأقل من نسبة زيادة الدخل؛ أي أن $1 > \beta_2 > 0$، ونستطيع اثبات أن الاستهلاك يعتمد على الدخل بإجراء معاكس للفرضية الأساسية التي تقول بأن الاستهلاك لا يعتمد على الدخل؛ حيث أن $H_0 : \beta_2 = 0$، والفرضية البديلة هي $H_1 : \beta_2 \neq 0$ حيث أن الدخل يؤثر على الاستهلاك، فإذا رفضت الفرضية الأساسية تكون العلاقة مثبتة على الأقل بمفهوم عام.

Dependent Variable: Y
Method: Least Squares
Sample: 1976 2007
Included observations: 32

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-273.4160	104.3785	-2.619467	0.0137
X	0.749767	0.018167	41.26971	0.0000

R-squared	0.982691	Mean dependent var	3247.813
F-statistic	1703.189	Durbin-Watson stat	0.534406

تبيّن نتائج تقدير انحدار الاستهلاك الفردي على الناتج المحلي الإجمالي للأردن بأن أول عامودين يبيّنا أسماء المتغيّرات وتقدير معاملاتها، والعمود الثالث يبيّن الخطأ المعياري لها، واحصائية t للفرضية الأساسية $H_0 : \beta_2 = 0$، وباستخدام (2.56) **يتم قسمة قيمة المعلمة المقدّرة على الخطأ المعياري لها للحصول على قيمة t المحسوبة:**

$$t = \frac{b_2 - \beta_2^\circ}{s.e.(b_2)} = \frac{b_2 - 0}{s.e.(b_2)} = \frac{0.749767}{0.018167} = 41.26971 \qquad (2.56)$$

وبما أن عدد المشاهدات 33 مشاهدة في العينة، ولدينا تقدير لمعلمتين تكون درجات الحرية (31)، وتكون القيمة الحرجة لـ (30) مشاهدة وهي الأقرب إلى (31) عند مستوى معنوية 5٪ هي (2.042) وبالتالي نتأكد من أننا سنرفض H_0 عند مستوى معنوية 5٪ لـ 31 درجة حرية، ونستنتج أن الدخل يؤثر على الاستهلاك الفردي.

بالطبع بما أننا استخدمنا مستوى معنوية 5٪ كأساس للاختبار، ستكون نسبة المخاطرة 5٪، وهي مخاطرة الخطأ من النوع الأول عندما تكون فرضية عدم الأثر صحيحة، وسنخفض الخطورة إلى 1٪ باستخدام مستوى معنوية 1٪ بدلاً من 5٪، وتكون قيمة t الحرجة عند مستوى معنوية 1٪ ودرجات حرية عددها 30 هي (2.750)، وبما أن قيمة t اكبر من هذه القيمة الحرجة سنرفض بكل سهولة الفرضية الأساسية عند هذا المستوى. لاحظ أنه عند اختبار 5٪ و 1٪ يؤدي إلى نفس النتيجة، فلا داعي لكتابة تقرير لكل منهما، ونكتفي بكتابة تقرير عند مستوى معنوية 1٪.

هذا الإجراء لبيان العلاقة بيّن المتغيّر التابع والمتغيّر التفسيري ودحض الفرضية الأساسية $H_0 : \beta_2 = 0$، علماً بأن جميع تطبيقات الانحدار المختلفة (مثل EViews و Stata و SPSS وغيرها) تنتج احصائية t وكحالة خاصة يتم قسمة قيمة المعلمة على الخطأ المعياري لها، وتعطي

النسبة قيمة إحصائية t، ومن نتائج تحليل دالة الاستهلاك تظهر احصائية t ومعلمة الميل في العمود المتوسط لنتائج الانحدار.

إذا حددت الفرضية الأساسية قيم غير صفرية للمعلمة β_2 يتم استخدام الصيغة العامة (2.51) وتحسب احصائية t يدوياً كما في مثال تضخم السعر/ تضخم الأجور.

Critical Values of the *t* Distribution

Significance Level					
1-Tailed: / 2-Tailed:		.10 / .20	.05 / .10	.025 / .050	.01 / .02
Degrees of Freedom	1	3.078	6.314	12.706	31.821
	2	1.886	2.920	4.303	6.965
	3	1.638	2.353	3.182	4.541
	4	1.533	2.132	2.776	3.747
	5	1.476	2.015	2.571	3.365
	:				
	16	1.337	1.746	2.120	2.583
	17	1.333	1.740	2.110	2.567
	18	1.330	1.734	2.101	2.552
	19	1.328	1.729	2.093	2.539
	20	1.325	1.725	2.086	2.528
	:				
	40	1.303	1.684	2.021	2.423
	60	1.296	1.671	2.000	2.390
	90	1.291	1.662	1.987	2.368
	120	1.289	1.658	1.980	2.358
	∞	1.282	1.645	1.960	2.326

2-9-1 - القيمة الاحتمالية

عندما نكتب تقريراً عن نتائج اختبار الفرضية الاحصائية، فقد أصبح من المعتاد كتابة القيمة الاحتمالية **p-value** للاختبار، وإذا توفر لدينا

القيمة الاحتمالية للاختبار نستطيع تحديد نتيجة الاختبار بمقارنة القيمة الاحتمالية بمستوى المعنوية المختار α دون النظر إلى/ أو حساب القيمة الحرجة.

القاعدة هي:

صندوق (2 -4) قاعدة القيمة الاحتمالية **p-value**
يتم رفض الفرضية الأساسية عندما تكون القيمة الاحتمالية أقل من/ أو تساوي مستوى المعنوية α، أي إذا كانت $P \leq \alpha$ نرفض H_0، أما إذا كانت $P > \alpha$ لا نستطيع رفض H_0.

إذا اخترنا مستوى المعنوية ليكون $\alpha = 0.01$ أو 0.05 أو 0.10 أو أي قيمة. نستطيع مقارنته بالقيمة الاحتمالية للاختبار ثم اتخاذ قرار الرفض أو عدمه دون فحص القيمة الاحتمالية. وفي كتابة تقرير العمل للقيمة الاحتمالية للاختبار نقبل الحكم بمستوى المعنوية المناسب.

يبيّن العمود الخامس من نتائج تحليل الانحدار (المعنوية .Prob) طريقة بديلة لصياغة معنوية معلمات الانحدار، ويبيّن الرقم في هذا العمود قيمة الاحتمالية P value لكل معامل. وتقابل هذه الاحتمالية إحصائية t. فإذا كانت الفرضية الأساسية للقيمة الاحتمالية أقل من 0.01 (1٪) فإنها تعني أن الفرضية الأساسية سترفض عند مستوى معنوية 1٪، والقيمة الاحتمالية بيّن 0.01 و 0.05 تعني أن الفرضية الأساسية سترفض عند مستوى معنوية 5٪ ولا ترفض عند 1٪، وإذا كانت القيمة الاحتمالية أكبر من 0.05 فإنها تعني عدم رفض الفرضية الأساسية عند مستوى معنوية 5٪.

2-9-2- فترات الثقة CONFIDENCE INTERVALS

النموذج النظري $Y_i = \beta_1 + \beta_2 X_i + u_i$ والنموذج المقدّر $\hat{Y}_i = b_1 + b_2 X_i$، ومعلمة الانحدار b_2 والقيمة الافتراضية $\beta_2 = \beta_2^{\circ}$ متعارضين إذا:

$$\frac{b_2 - \beta_2^{\circ}}{s.e.(b_2)} > t_{\text{الحرجة}} \quad \text{أو} \qquad (2.57)$$

$$\frac{b_2 - \beta_2^{\circ}}{s.e.(b_2)} < -t_{\text{الحرجة}}$$

$$b_2 - \beta_2^{\circ} > s.e.(b_2) \times t_{\text{الحرجة}} \quad \text{أو} \qquad (2.58)$$

$$b_2 - \beta_2^{\circ} < -s.e.(b_2) \times t_{\text{الحرجة}}$$

$$b_2 - s.e.(b_2) \times t_{\text{الحرجة}} > \beta_2^{\circ} \quad \text{أو} \qquad (2.59)$$

$$b_2 + s.e.(b_2) \times t_{\text{الحرجة}} < \beta_2^{\circ}$$

ويترتب على ذلك أن β_2 الإفتراضية متوافقة مع نتائج الانحدار إذا كان كل منهما:

$$\mathbf{b_2 + s.e.(b_2) \times t_{\text{الحرجة}} \geq \beta_2} \quad \text{و} \quad \mathbf{b_2 - s.e.(b_2) \times t_{\text{الحرجة}} \leq \beta_2} \qquad (2.60)$$

إذا حققت β_2 الإزدواج غير المتساوي:

$$\mathbf{b_2 - s.e.(b_2) \times t_{\text{الحرجة}} \leq \beta_2 \leq b_2 + s.e.(b_2) \times t_{\text{الحرجة}}} \qquad \mathbf{(2.61)}$$

كخلاصة لفترة الثقة، فأي قيمة افتراضية للمعلمة β_2 تحقق (2.61) ستتوافق مع تقدير b_2 ولا ترفض، ولعمل فترة ثقة نحتاج إلى اختبار مستوى معنوية وتحديد القيمة الحرجة لاحصائية t المقابلة.

مثال

كان معامل GDP في نتائج دالة الاستهلاك 0.749767، وكان الانحراف المعياري له 0.018167 والقيمة الحرجة لاحصائية t عند مستوى معنوية 5٪ كانت 2.042، وقيمة فترة الثقة المقابلة عند 95٪ هي:

$$0.74977 - 0.01817 \times 2.042 \leq \beta_2 \leq 0.74977 + 0.01817 \times 2.042$$

$$0.713 \leq \beta_2 \leq 0.787$$

لذا سنرفض القيمة الافتراضية عندما تكون أقل من 0.713 واكبر من 0.787؛ فأي فرضية داخل هذا التحديد لا ترفض.

تمارين

6-2 افترض الباحث أن الاستهلاك CONS قد يرتبط بمستوى الأسعار حسب العلاقة التالية:

$$CONS = \beta_1 + \beta_2 P + u$$

لاختبار الفرضية الأساسية $H_0: \beta_2 = 0$ مقابل الفرضية البديلة $H_1: \beta_2 \neq 0$ عند مستوى معنوية 5٪ و 1٪ لعينة من 60 مشاهدة. ماذا ستكتب:

1- إذا كان $b_2 = -0.20$ و $s.e.(b_2) = 0.07$؟

2- إذا كان $b_2 = -0.12$ و $s.e.(b_2) = 0.07$؟

3- إذا كان $b_2 = 0.06$ و $s.e.(b_2) = 0.07$؟

4- إذا كان $b_2 = 0.20$ و $s.e.(b_2) = 0.07$؟

7-2 في تمرين (2-1) انحدار معدل نمو البطالة على معدل نمو الناتج المحلي الإجمالي لعينة مكوّنة من 25 دولة من دول OECD، طبّق اختبار t على معامل الميل والحد الثابت وحدد نتيجتك.

8-2 احسب فترة ثقة 99٪ لـ b_2 في دالة الاستهلاك، حيث $b_2 = 0.749767$ وانحراف معياري $s.e.(b_2) = 0.018167$.

9-2 احسب فترة ثقة 95٪ لـ b_2 في مثال تضخم السعر/تضخم الأجور:

$$\hat{P} = \underset{(0.05)}{-1.21} + \underset{(0.10)}{0.82}\, w$$

وماذا تستنتج من هذه النتيجة؟

2-10- جودة التقدير R^2 GOODNESS OF FIT

يوجد سببان لتحليل نموذج الانحدار التالي:

$$Y_i = \beta_1 + \beta_2 X_i + u_i \qquad (2.62)$$

هما شرح كيفية تغير المتغيّر التابع Y_i عندما يتغيّر المتغيّر المستقل X_i، والتوقع عند تحديد X_i، واستخدام X_i لشرح التغيّرات في المتغير التابع Y_i إذا أمكن، ويسمى X_i في نموذج الانحدار (2.62) بالمتغير التفسيري لأننا نأمل أن تشرح تغيّراته أو تفسر التغيّرات في Y_i. ولتطوير مقياس التغيّرات في Y_i التي يفسرها النموذج سنبدأ بتجزئة Y_i إلى مكوّنين: مكوّن مُفَسَّر ومكوّن غير مُفَسَّر، وسنفترض أن:

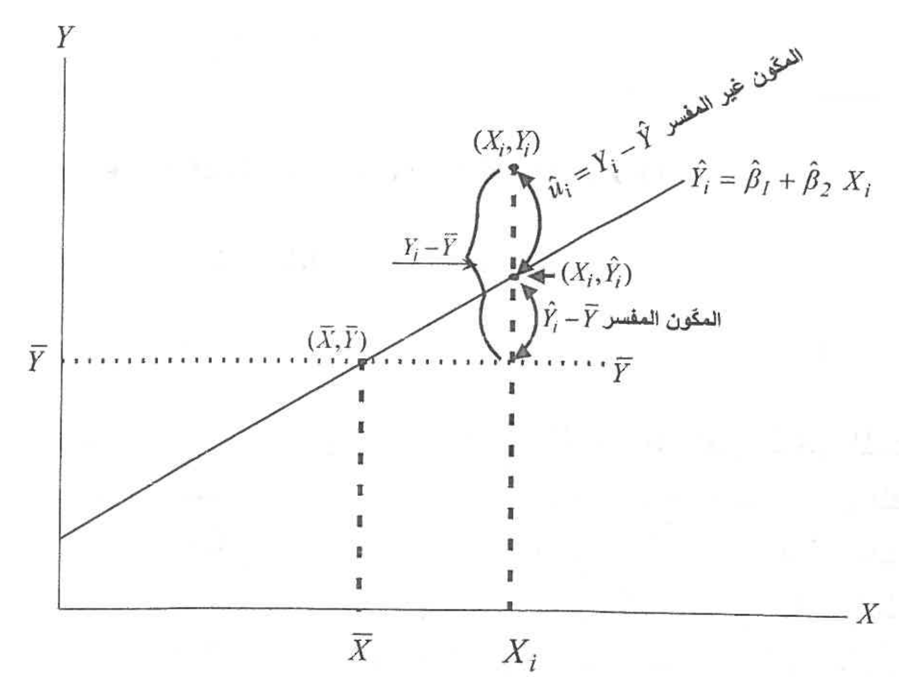
$$Y_i = E(Y_i) + u_i \qquad (2.63)$$

حيث أن $E(Y_i) = \beta_1 + \beta_2 X_i$ مكوِّن مُفَسَّر للمتغيّر Y_i، و u_i عشوائي ومكوِّن غير مُفَسَّر للمتغيّر Y_i، وكل منهما غير مشاهد، إلا أننا نستطيع تقدير المعلمات β_1 و β_2 المجهولة وتحليل قيمة Y_i إلى:

$$Y_i = \hat{Y}_i + \hat{u}_i \qquad (2.64)$$

حيث أن $\hat{Y}_i = b_1 + b_2 X_i$ و $\hat{u}_i = Y_i - \hat{Y}_i$. ويظهر الشكل (2-6) أن نقطة متوسط $(\bar{X}, \bar{Y})$ تخترق خط المربعات الصغرى المقدَّر، وهذه خاصية لخط المربعات الصغرى المقدر عندما يتضمن نموذج الانحدار الحد الثابت intercept، وبطرح متوسط العينة $\bar{Y}$ من كلا جانبي المعادلة نحصل على:

$$(Y_i - \bar{Y}) = (\hat{Y}_i - \bar{Y}) + \hat{u}_i \qquad (2.65)$$



شكل (2-6) مكوّنات Y المفسّرة وغير المفسّرة

كما يبدو من الشكل (2-6) أن الفرق بين Y_i ومتوسطها $\bar{Y}$ يتكون من جزءٍ "مُفَسَّر **Explained**" وجزءٍ "**غير مُفَسَّرٍ Unexplained**". وإذا تم تربيع كلا الجانبين نحصل على:

$$\sum(Y_i-\bar{Y})^2=\sum(\hat{Y}_i-\bar{Y})^2+\sum\hat{u}_i^2 \qquad (2.66)$$

تعطينا المعادلة (2.66) تجزئة تغيّرات العينة الكلية إلى مكوِّنات مُفَسَّرة ومكوِّنات غير مُفَسَّرة، وعلى وجه التحديد مجموع المربعات التي هي:

1- $\sum(Y_i-\bar{Y})^2$= مجموع المربعات الكلي = Total Sum of Squared = SST، ويقيس مجموع الانحرافات عن متوسط Y.

2- $\sum(\hat{Y}_i-\bar{Y})^2$ = مجموع المربعات نتيجة الانحدار = Explained SSE = Sum of Squared، وهو جزء من الانحرافات الكلية عن متوسط عينة Y المُفَسَّر بالانحدار، ويُعرف كذلك "مجموع المربعات المُفَسَّرة Explained Sum of Squared".

3- $\sum\hat{u}_i^2$ = مجموع مربعات الأخطاء = SSR، وهي الجزء من الانحرافات الكلية عن المتوسط التي لم يُفسِّرها الانحدار، وتسمى كذلك مجموع المربعات غير المُفَسَّرة أو مجموع مربع البواقي أو مجموع مربعات الأخطاء Sum of Squared Residuals.

وباستخدام المختصرات نحصل على:

$$SST=SSE+SSR \qquad (2.67)$$

وتحليل التغيّرات الكلية إلى جزء مُفَسَّر بنموذج الانحدار وجزء غير مُفَسَّر يسمح لنا بتعريف مقياس يسمى **معامل التحديد** Coefficint of determination أو R^2؛ وهو نسبة التغيّر في Y المفسرة بالمتغيّر X في نموذج الانحدار.

$$R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST} \qquad (2.68)$$

يقترب R^2 من 1، فإذا كان $R^2 = 1$ ستقع عينة البيانات كلها على خط الانحدار المقدر بالضبط وتكون $SSR = 0$، أما إذا لم يوجد ارتباط بين بيانات X و Y سيكون خط المربعات الصغرى المقدرة أفقياً ويماثل $\bar{Y}$ وبالتالي يكون $SSE = 0$ و $R^2 = 0$، وعندما تكون $0 < R^2 < 1$ تُفسِّر بنسبة انحرافات Y عن متوسطها التي تفسر بنموذج الانحدار، وتبين قيمة R^2 المنخفضة أهمية المتغيّرات المفقودة من النموذج، وكذلك الخصائص غير المشاهدة مهمّة في تحديد المتغيّر التابع، ونادراً ما تكون R^2 أكبر من 0.50 حتى في النموذج المحدد تماماً.

في ضوء (2.67) فإن $\sum(\hat{Y}_i - \bar{Y})^2 / \sum(Y_i - \bar{Y})^2$ نسبة مجموع المربعات المفسرة الكلية على خط الانحدار تسمى بمعامل التحديد Coefficint of determination:

$$R^2 = \frac{SSE}{SST} = \frac{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} \qquad (2.69)$$

من السهولة أن نرى معياراً مكافئ، وفي ضوء (2.67) نعيد كتابة R^2 كما يلي:

$$R^2 = \frac{SST - SSR}{SST} = 1 - \frac{SSR}{SST} = 1 - \frac{\sum_{i=1}^{n} u_i^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} \qquad (2.70)$$

وبالتالي تخفض قيم b_1 و b_2 مجموع مربعات البواقي وتعظم R^2 تلقائياً.

مثال: احسب معامل التحديد من بيانات الجدول أدناه.

الحل: نتبع الخطوات التالية ونحصل على:

X	Y	$\hat{Y}$	$\hat{Y}-\bar{\hat{Y}}$	$(\hat{Y}-\bar{\hat{Y}})^2$	$Y-\bar{Y}$	$(Y-\bar{Y})^2$
20	30	42.6316	-2.3684	5.6093	-15	225
40	60	66.3156	21.3156	454.3548	15	225
20	40	42.6316	-2.3684	5.6093	-5	25
30	60	54.4736	9.4736	89.7491	15	225
10	30	30.7896	-14.2104	201.9355	-15	225
10	40	30.7896	-14.2104	201.9355	-5	25
20	40	42.6316	-2.3684	5.6093	-5	25
20	50	42.6316	-2.3684	5.6093	5	25
20	30	42.6316	-2.3684	5.6093	-15	225
30	70	54.4736	9.4736	89.7491	25	625
220	**450**	**450**	0	**1065.7705**	0	**1850**
22	**45**	**45**		**SSE**		**SST**

تظهر قيمة R^2 دائماً كجزء من نتائج الانحدار التي يقدرها الحاسوب، وهذا المثال للتوضيح فقط، وحيث أظهرت نتائج المثال الذي بدأنا به في هذا الفصل خط الانحدار كما يلي:

$$\hat{Y}_i = 18.9476 + 1.1842\ X_i$$

حُسبت $\hat{Y}_i$ و u_i من الجدول أعلاه لكل مشاهدة، وتم حساب مجموع قيم $\sum(Y_i-\bar{Y})^2 = 1850$ و $\sum(\hat{Y}-\bar{\hat{Y}})^2 = 1\ 0\ .6\ 5$ و $\sum u^2 = 784.2105$، ومن هذه الأرقام نستطيع حساب R^2 :

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{Y}_i - \overline{Y})^2}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2} = \frac{1065.7705}{1850} = 0.576$$

ونحصل على نفس النتيجة باستخدام القانون التالي:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} u_i^2}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2} = 1 - \frac{784.2105}{1850} = 0.576$$

تمرين (2) تقدير معامل التحديد

X	Y	$X_i - \overline{X}$	$Y_i - \overline{Y}$	$(X_i - \overline{X})^2$	$(X_i - \overline{X})(Y_i - \overline{Y})$
5	160	-7.5	-77.5	56.25	581.25
5	220	-7.5	-17.5	56.25	131.25
5	140	-7.5	-97.5	56.25	731.25
10	190	-2.5	-47.5	6.25	118.75
10	240	-2.5	2.5	6.25	-6.25
10	260	-2.5	22.5	6.25	-56.25
15	230	2.5	-7.5	6.25	-18.75
15	270	2.5	32.5	6.25	81.25
15	280	2.5	42.5	6.25	106.25
20	260	7.5	22.5	56.25	168.75
20	290	7.5	52.5	56.25	393.75
20	310	7.5	72.5	56.25	543.75
		0	0	375	2775.00
12.5	237.5				

تم استخدام البيانات أعلاه لتقدير المعادلة $\hat{Y} = 145 + 7.4\ X_i$ كما في المثال السابق.

المطلوب:

ا- قدّر قيمة معامل التحديد R^2

2- فسر النتيجة.

X	Y	$\hat{Y}$	$\left(Y_i - \bar{Y}\right)^2$	$\hat{Y}_i - \bar{Y}$	$\left(\hat{Y}_i - \bar{Y}\right)^2$
5	160	$\hat{Y}_1 = 145 + 7.4 \times 5 = 182$			
5	220				
5	140				
10	190				
10	240				
10	260				
15	230				
15	270				
15	280				
20	260				
20	290				
20	310				

2-11- اختبار F

رأينا أن فروقات المتغيّر التابع قد تتحلل إلى مكوّن "مفسَّر" ومكوّن "غير مفسَّر" باستخدام المعادلة (2.66).

$$\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^{n}u_i^2 \qquad \textbf{(2.71)}$$

الجانب الأيسر هو مجموع المربعات الكلي (SST) لقيم المتغيّر التابع حول متوسطها، والحد الأول من الجانب الأيمن هو مجموع المربعات (SSE) "المفسَّر"، والحد الثاني هو مجموع مربع البواقي (SSR) "غير المفسَّر"؛ أي أن:

$$SST = SSE + SSR \qquad (2.72)$$

تكتب احصائية F لاختبار أحسن تقدير للانحدار كما يلي:" مجموع المربعات المفسَّر لكل متغيّر تفسيري بدرجات حرية $(k-1)$ مقسوماً على مجموع مربع البواقي لكل درجات الحرية الباقية $(n-k)$".

$$F(k-1,\ n-k) = \frac{SSE/(k-1)}{SSR/(n-k)} = \frac{\dfrac{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2}{(k-1)}}{\dfrac{\sum_{i=1}^{n} u_i^2}{(n-k)}} \qquad (2.73)$$

حيث k عدد معلمات معادلة الانحدار (المقطع أو الحد الثابت)، و $k-1$ معامل ميل.

يتم مقارنة قيمة F المحسوبة مع قيمة F الحرجة ($F_{\text{الحرجة}}$) في الجدول (2)، فإذا كانت F المحسوبة أكبر من $F_{\text{الحرجة}}$ ترفض الفرضة الأساسية $H_0 : \beta_2 = 0$ ونستنتج أن "المفسَّر" من Y أفضل مما يظهر بالصدفة، وعادة ما تظهر F في نتائج الانحدار.

يبيّن جدول (2) في الملحق المستويات الحرجة لـ F عند مستوى معنوية 1٪ و 5٪ و 10٪، وفي كل حالة يعتمد مستوى المعنوية على عدد المتغيّرات التفسيرية $k-1$ التي تقرأ من الجهة العليا للجدول، وعدد درجات الحرية $n-k$ التي تقرأ من الجانب الأيسر، وبالنسبة للانحدار البسيط تكون k تساوي 2 ونستخدم العمود الأول من الجدول.

يشبه هذا الاختبار اختبار t للمعلمات، ولا يكون خالياً من الأخطاء، وقد نجعل درجة الخطورة عند مستوى معنوية 5٪، ويكون الخطأ من النوع الأول (نرفض الفرضية الأساسية عندما تكون صحيحة في الواقع) بنسبة 5٪، وبالطبع قد نخفض الخطورة باستخدام مستوى معنوية أدق مثل مستوى 1٪، وسيتجاوز المستوى الحرج لـ F نسبة 1٪ إذا كانت H_0 صحيحة. وتكون اكبر من المستوى الحرج لاختبار بنسبة 5٪.

يبيّن الجدول (2-3) بما يُعرف بتحليل التباين **Analysis of Variance (ANOVA)**، ويتضمن الفروقات الكلية **Total variation** في **Y** والفروقات المُفسّرة بالمتغيّر **X** والفروقات غير المُفَسّرة. ويبيّن كذلك نسبة المُفَسّر إلى غير المُفَسّر التي تعطينا اختبار معنوية العلاقة الكلية، وهذا الفصل يتضمن متغيّر تفسيري واحد فقط، وتكافيء هذه العلاقة اختبار t للفرضية $H_0 : \beta_2 = 0$، وهي إحصائية اختبار F التي تساوي مجموع مربع قيمة إحصائية t، وبالتالي سنقدّر العلاقة في مثالنا حول X و Y التي تعطينا جدول تحليل التباين.

المعلومات في عمود الفروقات Variation هي: مجموع المربعات المُفَسَّرة، وغير المُفَسَّرة أو مجموع مربع الأخطاء، ومجموع المربعات الكلي على التوالي. ودرجات الحرية هي 1 و (N-2) و (N-1). وكل "مربع متوسط" نتج عن قسمة كل مصدر اختلاف على رقم درجات الحرية المقابل له، وبالتالي تكون 1065.7705 =1065.7705/1 و 784.2105/8=98.0263، وأخيراً قيمة F هي نسبة متوسط المربعين $F = 1065.7705/98.0263 = 10.872$، وهي تساوي (قيمة t)^2؛ أي أن قيمة $F = 10.872 = (3.297345)^2$.

جدول (2-3) تحليل التباين

	مصدر الاختلاف	درجات الحرية	متوسط مجموع المربعات	المفسر / غير المفسر
المُفَسّر	1065.7705	1	1065.7705/1= 1065.7705	**1065.7705/98.026 = 10.8723**
غير المُفَسّر	784.2105	8	784.2105/8=98.026	
الكلي	1850	9	205.5555	
بشكل عام				
المُفَسّر SSE	$\sum_{i=1}^{n}(\hat{Y}_i-\bar{Y})^2$	1	$\frac{\sum_{i=1}^{n}(\hat{Y}_i-\bar{Y})^2}{1}$	$F=\frac{SSE/(k-1)}{SSR/(n-k)}$
غير المُفَسّر SSR	$\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2$	N-2	$\frac{\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2}{(N-2)}=\sigma^2$	
الكلي SST	$\sum_{i=1}^{n}(Y_i-\bar{Y})^2$	N-1	$\frac{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}{N-1}$	

مثال

قُدر انحدار وكان مجموع المربعات المفسرة $SSE=19322$ و مجموع مربع البواقي غير المفسرة $SSR=92.689$ و مجموع المربعات الكلي $SST=112010$ وعدد المشاهدات 540 و $k=2$ وبالتالي يكون عدد درجات الحرية 538.

$$F = \frac{SSE/(k-1)}{SSR/(n-k)} = \frac{19322/1}{92689/538} = \frac{19322}{172.28} = 112.15$$

إذا كانت $H_0 : \beta_2 = 0$ صحيحة، فإنه لا يوجد علاقة حقيقية، وأنظر إلى الجدول (2) عند مستوى معنوية 0.01 والمستوى الحرج لـ F عند درجة حرية 1 و 500 (العمود الأول والسطر500) تساوي 10.96، عندها لا تتردد في رفض الفرضية الأساسية في هذا المثال.

تمرين (3) اختبار F

1- استخدم بيانات تمرين (2) لحساب إحصائية F

2- اختبر معنوية F المحسوبة

X	Y	$\hat{Y}$	$Y-\hat{Y}_i$	$(Y-\hat{Y}_i)^2$	$\hat{Y}_i-\overline{Y}$	$(\hat{Y}_i-\overline{Y})^2$
5	160					
5	220					
5	140					
10	190					
10	240					
10	260					
15	230					
15	270					
15	280					
20	260					
20	290					
20	310					

2-11-1- العلاقة بين اختبار F واختبار T لمعامل الميل في تحليل الانحدار البسيط

في سياق تحليل الانحدار البسيط (فقط في تحليل الانحدار البسيط)، فإن اختبار F واختبار t بذيلين لمعامل الميل لهما نفس الفرضية الأساسية $H_0 : \beta_2 = 0$ ونفس الفرضية البديلة $H_1 : \beta_2 \neq 0$، وتساوي احصائية F مربع إحصائية t، وعند أي مستوى معنوية تساوي القيمة الحرجة لـ F مربع قيمة t الحرجة، ونبدأ من تعريف F في (2.74) و $k = 2$.

$$F = t^2 \qquad (2.74)$$

2-12- التنبؤ PREDICTION

إن القدرة على التنبؤ Prediction مهمّة لاقتصاديي الأعمال والمحلّلين الماليين لتوقع مبيعات وإيرادات شركة ما، ومهم لصانعي السياسة الحكومية الذين يحاولون التنبؤ بمعدلات نمو الدخل القومي، والتضخم، والاستثمار والادخار، ونفقات الضمان الاجتماعي، وإيرادات الضرائب، وكذلك مهم لرجال الأعمال المحليين للتنبؤ بنمو السكان والدخل من أجل توسيع أو تركيز خدماتهم، ويعتبر التنبؤ الدقيق أساساً لصناعة قرار أفضل، وفي هذا الجزء سنستكشف استخدام الانحدار الخطي كأداة للتنبؤ.

خذ الانحدار الخطي البسيط وفرضياته، وافرض أن X_0 قيمة المتغيّر التفسيري، ونريد التنبؤ بقيمة Y المسماة Y_0، ومن أجل استخدام تحليل

الانحدار كأساس للتنبؤ، يجب أن نفترض أن Y_0 و X_0 مرتبطان ببعضهما في نموذج الانحدار الذي يصف عينة البيانات.

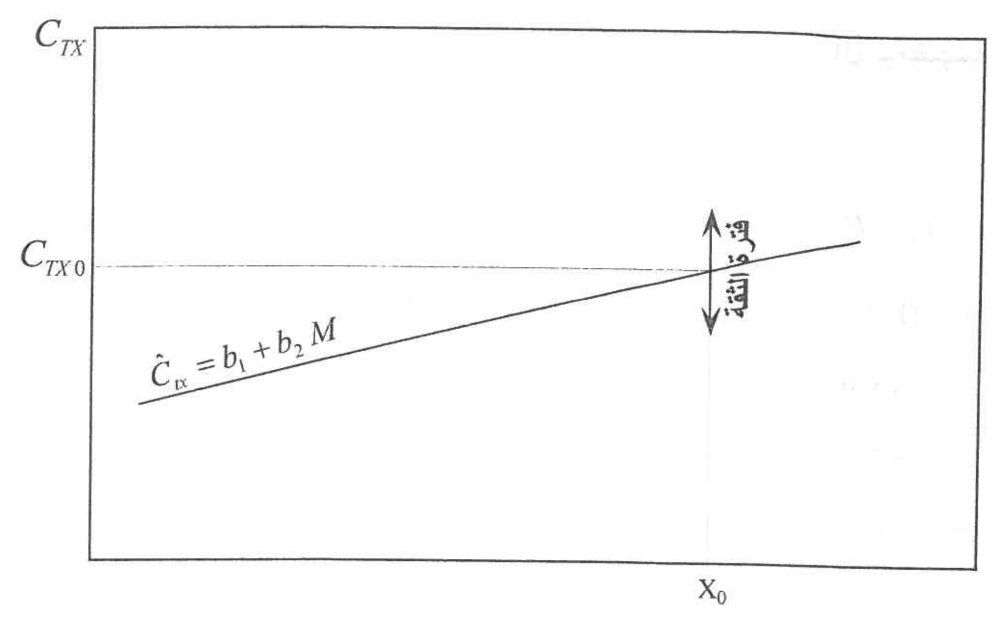
$$Y_0 = \beta_1 + \beta_2 X_0 + u_0 \qquad (2.75)$$

حيث أن u_0 الخطأ العشوائي، ونفترض أن $E(Y_0) = \beta_1 + \beta_2 X_0$ و $E(u_0) = 0$، ونفترض كذلك أن u_0 لها نفس التباين $\text{var}(u_0) = \sigma^2$، و u_0 غير مرتبطة بالأخطاء العشوائية وبالتالي $\text{cov}(u_0, u_i) = 0$ حيث أن $i = 1,2,3,\cdots,N$، ونحصل على تنبؤ المربعات الصغرى Y_0 من خط الانحدار المقدر كما يلي:

$$\hat{Y}_0 = b_1 + b_2 X_0 \qquad (2.76)$$

تُعطى القيمة المتنبأ بها $\hat{Y}_0$ بنقطة على خط المربعات الصغرى المقدر، حيث $X = X_0$، كما في الشكل (2-7)، لكن ما هو الإجراء الأفضل في التنبؤ؟ نتبع مئينات المربعات الصغرى $\hat{Y}_0 = b_1 + b_2 X_0$، ولتقييم كيفية تطبيق التنبؤ سنعرّف أخطاء التوقع **Forecast error** المشابهة (المماثلة) لبواقي المربعات الصغرى كما يلي:

$$\begin{aligned} E(f) &= E(\beta_1 + \beta_2 X_0 + u_0) - E(b_1 + b_2 X_0) \\ &= \beta_1 + \beta_2 X_0 + E(u_0) - E(b_1) - X_0\ E(b_2) \\ &= \beta_1 + \beta_2 X_0 - 0 - \beta_1 - X_0\ \beta_2 \\ &= 0 \end{aligned}$$



شكل رقم (2-7) ثبات تباين البيانات

الذي يعني أن الوسط الحسابي لخطأ التوقع (التنبؤ) يساوي صفر، وأن $\hat{Y}_0$ هي تنبؤ غير منحاز unbiased predictor للمتغيّر Y_0. وعلى كل حال، فإن عدم التحيّز ليس شرطاً ضرورياً؛ مما يعني أن التوقع مطابق للقيمة الفعلية، وأن احتمال أخطاء التنبؤ يعتمد على تباين خطأ التنبؤ، وبالتالي فإن $\hat{Y}_0$ هي أفضل خط تنبؤ غير منحاز **best linear unbiased predictor (BLUP)** للمتغيّر Y_0 إذا توفرت الفرضيات من الفرضية 1 إلى 6، وهذه النتيجة تعطي معلمات المربعات الصغرى b_1 و b_2 وتكون أفضل تقدير خطي غير متحيز.

نستطيع أن نرى من التباين-التباين المشترك لمعلمات المربعات الصغرى أن تباين أخطاء التوقع كما يلي:

$$\operatorname{var}(f) = \sigma_f^2 = \sigma_u^2 \left\{ 1 + \frac{1}{N} + \frac{\left(X_0 - \overline{X}\right)^2}{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2} \right\} \quad (2.77)$$

يفضل أن يكون تباين أخطاء التوقع صغيراً وتزيد احتمالية توقع $\hat{Y}_0$ للتطابق مع قيمة Y_0 التي نحاول توقعها، مع ملاحظة أن تباين أخطاء التوقع تكون أصغر عندما يكون:

أ) جميع المخاطر في النموذج صغيرة وتقاس بتباين الأخطاء العشوائية σ^2.

ب) حجم العينة كبيراً N .

ج) تنوع كبير في المتغيّر العشوائي.

د) قيمة $\left(X_0 - \overline{X}\right)^2$ صغيرة.

الإضافة الجديدة هي الحد $\left(X_0 - \overline{X}\right)^2$ الذي يقيس مسافة بُعد X_0 عن مركز قيم X ، وزيادة بُعد X_0 عن مركز بيانات العينة تجعل تباين التوقع كبيراً، وهذا يعني أننا سنكون قادرين على إجراء تنبؤ أفضل، حيث يكون لدينا المزيد من المعلومات، ويكون لدينا دقة تنبؤ أقل عندما نحاول التنبؤ خارج حدود البيانات.

نستبدل في التطبيق العملي σ^2 في المعادلة (2.77) بتقدير $\hat{\sigma}^2$ للحصول على:

$$\hat{\text{var}}(f) = \hat{\sigma}_f^2 = \hat{\sigma}_u^2 \left\{ 1 + \frac{1}{N} + \frac{\left(X_0 - \bar{X}\right)^2}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2} \right\} \qquad (2.78)$$

نأخذ الجذر التربيعي للتباين المقدر ونحصل على الخطأ المعياري للتوقع:

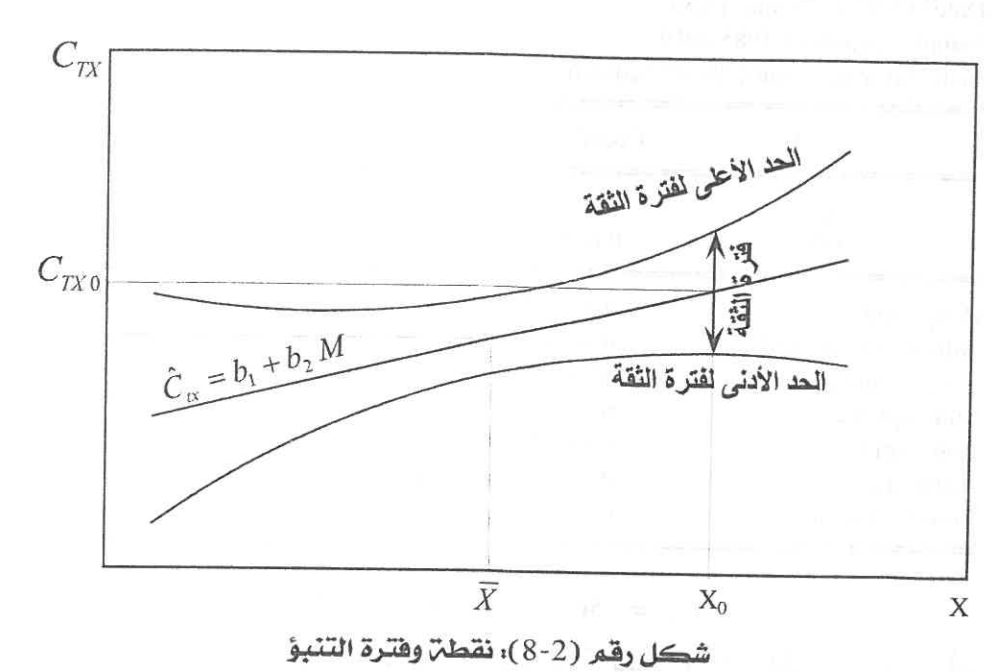
$$\textbf{s.e.}(f) = \sqrt{\hat{\text{var}}(f)} = \sqrt{s_u^2 \left\{ 1 + \frac{1}{n} + \frac{\left(X_0 - \bar{X}\right)^2}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2} \right\}} \qquad \textbf{(2.79)}$$

تعريف القيمة الحرجة $t_{\text{الحرجة}}$ ليكون الاحتمال $100(1-\alpha/2)$ من توزيع t، نستطيع الحصول على فترة التنبؤ باحتمال $100(1-\alpha)$ كما يلي:

$$\hat{Y}_0 \pm t_{\text{الحرجة}}\, se(f) \qquad (2.80)$$

وفيما يلي بعض التفاصيل المتعلقة بالتباين $\text{var}(f)$ في المعادلة (2.77)، فإذا كان بُعد X_0 عن وسط العينة $\bar{X}$ كبيراً سيكون تباين خطأ التنبؤ كبيراً، وتكون الثقة بالتنبؤ أقل؛ وبعبارة أخرى، فإذا كان التنبؤ بقيم X_0 يطابق وسط العينة $\bar{X}$ يكون أكثر ثقة من التنبؤ في حالة ابتعاد قيم X_0 عن وسط العينة $\bar{X}$، وتظهر هذه الحقيقة في حجم فترة التنبؤ،

والعلاقة بين نقطة وفترة التنبؤ بقيم X_0 المختلفة التي يشرحها الشكل (2-8)، وتعطى نقطة التنبؤ بخط المربعات الصغرى المقدر $\hat{Y}_0 = b_1 + b_2 X_0$ ، وتعطي فترة التنبؤ شكل نطاق (حزام) حول خط المربعات الصغرى المقدرة، ولأن تباين التوقع يتزايد ويبتعد X_0 عن وسط العينة $\bar{X}$ ، سيكون نطاق الثقة ضيقاً عندما $X_0 = \bar{X}$ ويتزايد العرض عندما يتزايد $|X_0 - \bar{X}|$.



شكل رقم (2-8): نقطة وفترة التنبؤ

2-12-1- التنبؤ في نموذج تقدير الضرائب الجمركية

يمكن استخدام المعادلة المقدرة في التنبؤ prediction أو التوقع forecasting، وافرض أننا نريد التنبؤ بالايرادات الجمركية (الضرائب الجمركية) المسماة بالرسوم الجمركية $\hat{C}_{tx}$ بالاعتماد على حجم المستوردات M حسب النموذج التالي:

$$\hat{C}_{tx} = b_1 + b_2 M$$

تم تقدير المعادلة أعلاه باستخدام بيانات السلسلة الزمنية للفترة 1985-2010 بعد استثناء مستوردات النفط ومشتقاته من قيمة المستوردات السلعية الإجمالية، وكانت النتائج كما يلي:

Dependent Variable: CUSTOMSTAX
Method: Least Squares
Date: 12/07/15 Time: 13:59
Sample (adjusted): 1985 2010
Included observations: 26 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	156.0288	17.15316	9.096211	0.0000
MN	0.018889	0.003995	4.727767	0.0001

R-squared	0.482221	Mean dependent var	220.6269
Adjusted R-squared	0.460646	S.D. dependent var	71.99976
S.E. of regression	52.87712	Akaike info criterion	10.84762
Sum squared resid	67103.76	Schwarz criterion	10.94440
Log likelihood	-139.0191	Hannan-Quinn criter.	10.87549
F-statistic	22.35179	Durbin-Watson stat	0.577459
Prob(F-statistic)	0.000083		

$$\hat{C}_{tx} = 156.0288 + 0.0189\, M$$

وإذا أردنا التنبؤ بالايرادات الجمركية لعام 2011 على افتراض أن حجم المستوردات بدون نفط ومشتقاته هو 9000 مليون دينار، يتم تعويض قيمة المستوردات في المعادلة المقدرة التالية:

$$\hat{C}_{tx} 2011 = 156.0288 + 0.0189 \times 9000$$
$$= 326.0388$$

تم التنبؤ بالايرادات الجمركية لعام 2011 عندما كانت المستوردات من دون نفط ومشتقاته 9000 مليون دينار ستكون قيمة الايرادات الجمركية

326 مليون دينار. وسنكون قادرين على ربط "فترة الثقة" بهذا التنبؤ، حيث أن التباين المقدّر لخطأ التوقع هو:

$$\text{vâr}(f) = \hat{\sigma}_u^2 \left\{ 1 + \frac{1}{N} + \frac{\left(X_0 - \bar{X}\right)^2}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2} \right\}$$

$$= \hat{\sigma}_u^2 + \frac{\hat{\sigma}_u^2}{N} + \left(X_0 - \bar{X}\right)^2 \frac{\hat{\sigma}_u^2}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}$$

$$= \hat{\sigma}_u^2 + \frac{\hat{\sigma}_u^2}{N} + \left(X_0 - \bar{X}\right)^2 \text{vâr}(b_2) \qquad (2.81)$$

سيكون اهتمامنا في السطر الأخير لتقدير تباين b_2 من $\text{vâr}(b_2) = \hat{\sigma}_u^2 \Big/ \sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$، وحصلنا على قيمة $\hat{\sigma}_u^2 = 2795.9898$ و $\text{vâr}(b_2) = 0.0000$، وبيانات الضرائب الجمركية $N = 26$ ووسط عينة المتغيّر التفسيري $\bar{X} = 3419.915$، وباستخدام تلك القيم نحصل على الخطأ المعياري للتوقع $se(f) = \sqrt{\text{vâr}(f)} = \sqrt{2903.478} = 53.8839$، فإذا اخترنا $1 - \alpha = 0.95$ فإن $t_{\text{الحرجة}} = t_{(0.975,\,26)} = 2.052$ وفترة ثقة للمتغيّر Y_0 باحتمال 95٪ تساوي:

$$\hat{Y}_0 \pm t_{\text{الحرجة}}\, se(f) = 326.0388 \pm 2.052 \times 53.8839$$

$$= [215.469, \quad 436.6086]$$

تبين فترة التنبؤ أن 9000 مليون دينار ستؤدي إلى تحصيل ضرائب جمركية بمقدار 326 مليون دينار بين 215.5 و 436.6 مليون دينار، وحيث

أن الفترة واسعة، فإن هذا يعني أن نقطة التنبؤ 326 مليون دينار لا يمكن الاعتماد عليها للغاية، لأننا حصلنا على فترة تنبؤ واسعة للقيمة $X_0 = 9000$ وهي بعيدة عن وسط $\bar{X} = 3419.9$ ، وقيم X اكثر تطرفاً لفترة التنبؤ الواسعة، وحيث أن التنبؤ لا يمكن الاعتماد عليه قد يبرهن أن جمع عينة كبيرة من البيانات قد تبرهن على أن دقة تباين خطأ التقدير $\hat{\sigma}^2$ قريبة من تباين تقدير خطأ التوقع $\hat{\text{var}}(f)$ ، مبنية مبدئياً أن خطورة التنبؤ تأتي من زيادة الخطورة في النموذج، وهذا لا يثير دهشتنا، وحيث أن تنبؤ سلوك الضرائب ظاهرة معقدة على أساس خصائص الضرائب والمستوردات، ربما لأن المستوردات هي مفتاح الضرائب الجمركية إلا أننا نستطيع تخيل خصائص أخرى للضرائب قد تلعب دوراً. ولتحقيق دقة اكثر في تنبؤ الضرائب الجمركية قد نحتاج إلى تضمين معادلة الانحدار بعوامل أخرى مثل سعر الصرف.

تمارين

10-2 في تمرين (2-1) في انحدار معدل نمو العمل على معدل نمو GDP باستخدام عينة 25 دولة من دول OECD، حيث $SSE = 14.58$ و $SSR = 10.13$، احسب قيمة احصائية F وافحص فيما إذا كانت تساوي 33.1، وكذلك احسب احصائية F باستخدام $R^2 = 0.59$ وتحقق أنها نفسها. ونفذ اختبار F عند مستوى معنوية 5٪ و 1٪ و 10٪، ومن الضروري كتابة نتائج اختبار المستويات الثلاث جميعها.

11-2 فيما يلي 5 مشاهدات استخدمها لحساب التالي:

X	Y	$X_i - \bar{X}$	$(X_i - \bar{X})^2$	$Y_i - \bar{Y}$	$(X_i - \bar{X})(Y_i - \bar{Y})$
3	5				
2	2				
1	3				
-1	2				
0	-2				
$\sum X_i$	$\sum Y_i$	$\sum(X_i - \bar{X})$	$\sum(X_i - \bar{X})^2$	$\sum(Y_i - \bar{Y})$	$\sum(X_i - \bar{X})(Y_i - \bar{Y})$

أ) أكمل الجدول، واحسب المجموع في السطر الأخير، ومتوسط العينة $\bar{X}$ و $\bar{Y}$.

ب) احسب b_1 و b_2 وفسرها بالكلمات.

ج) احسب $\sum_{i=1}^{5} X_i^2$ و $\sum_{i=1}^{5} X_i Y_i$ واستخدم القيم الحسابية لبيان:

$$\sum (X_i - \bar{X})^2 = \sum X_i^2 - n\bar{X}^2$$

$$\sum (X_i - \bar{X})(Y_i - \bar{Y}) = \sum X_i Y_i - n\overline{XY}$$.

د) استخدم تقدير المربعات الصغرى في الجزء (ب) لحساب القيم المقدّرة لـ Y واكمل الجدول ادناه، واحسب المجموع في السطر الأخير.

X	Y	$\hat{Y}_i$	$\hat{e}_i$	$\hat{e}_i^2$	$X_i \hat{e}_i$
3	5				
2	2				
1	3				
-1	2				
0	-2				
$\sum X_i$	$\sum Y_i$	$\sum \hat{Y}_i$	$\sum \hat{e}_i$	$\sum \hat{e}_i^2$	$\sum X_i \hat{e}_i$

هـ) ارسم على ورقة رسم نقاط البيانات وحدد خط الانحدار المقدّر: $\hat{Y}_i = b_1 + b_2 X_i$

و) حدد نقاط الوسط الحسابي $\bar{X}$ و $\bar{Y}$ على الرسم في (هـ)، وهل الخط المقدّر يخترق تلك النقاط؟ إذا كان لا، أعد رسم الخط.

ز) بين القيم العددية $\bar{Y} = b_1 + b_2\ \bar{X}$.

ح) بين القيم العددية $\bar{\hat{Y}} = \bar{Y}$، حيث أن $\bar{\hat{Y}} = \sum \hat{Y}_i / n$.

ط) احسب $\hat{\sigma}^2$.

ي) احسب $v\hat{a}r(b_2)$.

2-12 فيما يلي بيانات تشير إلى الكمية المباعة من السلعة Y (مقاسة بالكيلوغرام) وسعرها X (مقاساً بالقرش/كغم) لعشرة أسواق مختلفة:

Y	198	181	170	179	163	145	167	203	251	147
X	23	24.5	24	27.2	27	24.4	24.7	22.1	21	25

أ) افترض أن العلاقة بين المتغيّرين خطية، وقدر انحدار المربعات الصغرى OLS للحصول على b_1 و b_2.

ب) ارسم خط انحدار OLS للعينة من خلال شكل انتشار للبيانات.

ج) قدّر مرونة الطلب لهذه السلعة عند نقطة وسط العينة (أي عند $Y = \bar{Y}$ و $X = \bar{X}$).

2-13 من دالة الاستهلاك الكينزية: $C_i = \alpha + \delta\, Y_i^d$

الميل الحدي للاستهلاك المقدّر هو δ بينما الميل المتوسط للاستهلاك هو $C/Y^d = \hat{\alpha}/Y^d + \hat{\delta}$، وباستخدام بيانات 200 عائلة سعودية عن

الدخل والاستهلاك (مقاسة بالريال السعودي) وجدنا معادلة الانحدار التالية:

$$C_i = 138.52 + 0.725\, Y_i^d, \quad R^2 = 0.862$$

أ) اشرح الحد الثابت في هذه المعادلة وعلق على اشارة الميل ومعناها.

ب) احسب قيمة الاستهلاك المتوقع لدخل عائلة سنوي افتراضي قدره 100000 ريال.

ج) عندما يكون Y^d على المحور السيني. ارسم الميل الحدي للاستهلاك MPC والميل المتوسط للاستهلاك APC المقدّرين.

14-2 احصل على بيانات عن معدل التضخم ومعدل البطالة لدولة ما.

أ) قدر معادلة الانحدار التالية المسماة بمنحنى فيليبس Phillips Curve:

$$\pi_t = a_0 + a_1\, UNEMP_{t-1} + u_t$$

حيث أن π_t التضخم و $UNEMP$ البطالة، ثم اعرض النتائج بالطريقة المعتادة.

ب) قدر النموذج البديل التالي:

$$\pi_t - \pi_{t-1} = a_0 + a_1\, UNEMP_{t-1} + u_t$$

ج) أعد تقدير المعاملات أعلاه بعد تجزئة البيانات إلى جزئين

مختلفين، وما هي المعاملات المحسوبة لهما؟ وأي فترة زمنية تكون فيها المعادلة افضل تقدير؟ وأوضح المعايير التي استخدمتها بتجزئة الفترتين.

15-2 فيما يلي معادلة مُقدَّرة بطريقة المربعات الصغرى OLS:

$$\hat{R}_t = \underset{(0.33)}{0.567} + \underset{(0.066)}{1.045}\, R_{mt}, \quad n = 250$$

حيث أن R_t و R_{mt} عائد السهم وعائد السوق لسوق الرياض المالي:

أ) هل تلك المعلمات معنوية إحصائياً؟ اشرح معنى نتائج المعادلة حسب نظرية تمويل رأس المال CAPM Theory.

ب) اختبر الفرضية $H_0: \beta = 1$ و $H_1: \beta < 1$ عند مستوى معنوية 1٪، فإذا رفضت H_0 ماذا يشير هذا السهم؟

16-2 احصل على بيانات عن التكوين الرأسمالي الثابت (الاستثمار I) وسعر الفائدة المناسب (r) وبالاعتماد على المعادلة التالية:

$$I_t = \alpha_0 + \alpha_1 r_t + e_t$$

أ) ماذا تتوقع أن تكون اشارة معاملات هذه المعادلة.

ب) اشرح ماذا تعني تلك الاشارات.

ج) كيف يمكن أن تستخدم هذه المعادلة لتقدير مرونة الاستثمار بالنسبة لسعر الفائدة؟

د) قدّر دالة الانحدار.

هـ) أي من المعلمات معنوية احصائياً؟ وهل الإشارات كما هو متوقع؟

و) قدر الشكل اللوغاريتمي الخطي log-Linear لدالة الانحدار التالية:

$$\ln I_t = a_0 + a_1 \ln r_t + u_t$$

ز) هل مرونة الاستثمار بالنسبة لسعر الفائدة المقدّرة ذات دلالة؟

ح) هل تتوقع أنها مرنة أم غير مرنة؟ ولماذا؟

ط) أكتب فرضية اختبار مرونة الاستثمار بالنسبة لسعر الفائدة.

ي) أكتب الملخص الاحصائي للمتغيّرات أعلاه واشرحها.

الفصل الثالث

نموذج الانحدار المتعدد

The Multiple Regression Model

شرحنا في الفصل السابق نموذج الانحدار البسيط الذي يفترض أن المتغيّر التابع يرتبط بمتغيّر تفسيري واحد، أما إذا كان لدينا عدة متغيّرات تفسيرية ونرغب بقياس أثر كل منها على المتغيّر التابع، سنستخدم أسلوب يعرف بـ "**تحليل الانحدار المتعدد Multiple Regression Analysis**" الذي نعالجه في هذا الفصل. وسيتم توسيع نموذج الانحدار البسيط، والتركيز على نموذج بمتغيّرين مستقلين اثنين، إضافة إلى نموذجٍ عام.

3-1- نموذج الانحدار بمتغيّرين تفسيريين

سنبدأ بمثال دالة الطلب على النقود في الأردن، وسيتم توسيع نموذج الانحدار البسيط إلى نموذج يسمح بتأثر الطلب على النقود (ممثلاً بعرض النقد M_1) بالناتج المحلي الإجمالي وسعر الفائدة، ونفترض العلاقة التالية:

$$M_1 = \beta_1 + \beta_2 GDP + \beta_3 R + u \qquad (3.1)$$

حيث أن M_1 عرض النقد الضيق في الأردن، و GDP الناتج المحلي الإجمالي، و R سعر الفائدة، و u حد الخطأ. وتعني المعادلة (3.1) رياضياً أنه إذا كان R يساوي صفراً، يكون الطلب على النقود يساوي

$\beta_1+\beta_2 GDP$ عند أي قيمة موجبة للناتج المحلي الإجمالي GDP، وعندما يزيد $\beta_2 GDP$ يكون الطلب على النقود هو "صافي تأثير الناتج المحلي الإجمالي GDP"، أما إذا كان GDP يساوي الصفر، فهذا يعني أنه عند أي قيمة موجبة لسعر الفائدة R سيكون الطلب على النقود يساوي $\beta_1+\beta_3 R$، وعندما يزيد $\beta_3 R$ يكون الطلب على النقود هو "صافي أثر سعر الفائدة" "Pure R effect"، وإذا مزجنا أثر الناتج المحلي الإجمالي وسعر الفائدة يصبح الأثر $\beta_2 GDP+\beta_3 R$.

جدول (3-1) المشاهدات الفصلية للناتج المحلي الاجمالي وعرض النقد الضيق والواسع وسعر الفائدة (القيمة: مليون دينار)

obs	GDP	M_1	M_2	R (%)
1992Q1	832.2	5013.1	11446.7	10.9
1992Q2	894.1	5191.1	11686.3	10.717
1992Q3	959.7	5429.8	12150.3	10.803
1992Q4	924.6	5357.1	12449.4	10.87
1993Q1	904.8	5136.7	12743.9	10.707
1993Q2	977.1	5315.5	13221.3	10.570
1993Q3	1038.3	5540.5	13653.0	10.673
1993Q4	964.1	5465.6	13675.0	10.657
1994Q1	960.8	5202.3	13592.3	10.663
⋮	⋮			⋮
2004Q1	1793.7	8654.1	28431.4	8.767
2004Q2	2036.6	8590.4	28724.9	8.4
2004Q3	2180.7	9318.4	30333.5	8.033
2004Q4	2153.4	9489.0	31347.2	7.833
2005Q1	2015.6	9764.8	32080.3	7.567
2005Q2	2272.7	10888.7	33693.5	7.433

الملخص الاحصائي

	GDP	M1	M2	R
Mean	1422.254	6058.839	19603.46	10.90189
Median	1384.65	5329	17567.25	10.885
Maximum	2272.7	10888.7	33693.5	12.803
Minimum	832.2	4660.4	11446.7	7.433
Std. Dev.	359.7448	1510.401	5983.731	1.42774
Skewness	0.455243	1.530971	0.632372	-0.69551
Kurtosis	2.519941	4.37084	2.306417	2.919351
Jarque-Bera	2.383743	25.32304	4.681426	4.36819
Probability	0.303652	0.000003	0.096259	0.11258
Sum	76801.7	327177.3	1058587	588.702
Sum Sq. Dev.	6859066	1.21E+08	1.90E+09	108.0374
Observations	54	54	54	54

ويتم تقدير المعادلة بعد أخذ اللوغاريتم الطبيعي لها وتصبح كما يلي:

$$\hat{m}_1 = b_1 + b_2\, gdp + b_3\, r \qquad (3.2)$$

حيث أن $m_1 = \log(M_1)$ و $gdp = \log(GDP)$ و $r = \log(R)$.

وتعتمد النتائج على خيارات b_1 و b_2 و b_3 كتقدير لـ β_1 و β_2 و β_3 على التوالي، وباستخدام بيانات البنك المركزي الأردني خلال الفترة

2005:02-1992:01 نحصل على تقدير نتائج الانحدار التالية: (قدّرت المعادلة بعد أخذ اللوغاريتمات الطبيعية لكل متغيّر)

Dependent Variable: LOG(M1)
Method: Least Squares
Date: 11/26/15 Time: 19:45
Sample: 1992Q1 2005Q2
Included observations: 54

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	8.665171	0.261507	33.13547	0.0000
LOG(GDP)	0.362177	0.025780	14.04867	0.0000
LOG(R)	-1.092359	0.046602	-23.44028	0.0000

R-squared	0.966707	Mean dependent var	8.683709
Adjusted R-squared	0.965401	S.D. dependent var	0.218661
S.E. of regression	0.040673	Akaike info criterion	-3.512568
Sum squared resid	0.084368	Schwarz criterion	-3.402068
Log likelihood	97.83932	Hannan-Quinn criter.	-3.469952
F-statistic	740.4168	Durbin-Watson stat	0.717899
Prob(F-statistic)	0.000000		

تفسر المعادلة كما يلي: زيادة الناتج المحلي الإجمالي بنسبة 1٪ مع بقاء سعر الفائدة ثابتاً يزيد الطلب على النقود بنسبة 0.362٪، وزيادة سعر الفائدة بنسبة 1٪ مع بقاء الناتج المحلي الإجمالي ثابتاً يُخفّض الطلب على النقود بنسبة 1.092٪، وعادة لا يكون للحد الثابت معنىً واضحاً.

3-2- اشتقاق وتفسير معاملات الانحدار المتعدد

نفترض حالة متغيّر تابع Y يتحدد بمتغيّرين تفسيريين اثنين هما X_2 و X_3، وتكون العلاقة الصحيحة لهما كما يلي:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \tag{3.3}$$

حيث u حد الخطأ، ويأخذ المتغيّر X حرفين منخفضين: يُشير الأول إلى تعريف المتغيّر X (الناتج المحلي الإجمالي، سعر الفائدة، ...)، ويُشير الثاني إلى المشاهدة، ويكتب النموذج المقدّر كما يلي:

$$\hat{Y}_i = b_1 + b_2 X_{2i} + b_3 X_{3i} \tag{3.4}$$

نختار قيم معاملات الانحدار التي تجعل التقدير جيد قدّر الامكان كما في حالة الانحدار البسيط، على أمل الحصول على تقديرٍ مُرْضٍ للمعلمات الصحيحة غير المعروفة، وكما سبق تعريفنا لأفضل تقدير يكون بتقليل مجموع مربعات البواقي $SSR = \sum u_i^2$، حيث u_i بواقي المشاهدات i، وهي الفرق بين القيمة الصحيحة Y_i للمشاهدات، وقيمة $\hat{Y}_i$ المقدّرة كما يلي:

$$u_i = Y_i - \hat{Y}_i = Y_i - b_1 - b_2 X_{2i} - b_3 X_{3i} \tag{3.5}$$

$$SSR = \sum u_i^2 = \sum (Y_i - b_1 - b_2 X_{2i} - b_3 X_{3i})^2$$

$$= \sum (Y_i^2 + b_1^2 + b_2^2 X_{2i}^2 + b_3^2 X_{3i}^2 - 2b_1 Y_i - 2b_2 X_{2i} Y_i$$

$$- 2b_3 X_{3i} Y_i + 2b_1 b_2 X_{2i} + 2b_1 b_3 X_{3i} + 2b_2 b_3 X_{2i} X_{3i})$$

$$= \sum Y_i^2 + nb_1^2 + b_2^2 \sum X_{2i}^2 + b_3^2 \sum X_{3i}^2 - 2b_1 \sum Y_i$$

$$- 2b_2 \sum X_{2i} Y_i - 2b_3 \sum X_{3i} Y_i + 2b_1 b_2 \sum X_{2i}$$

$$+ 2b_1 b_3 \sum X_{3i} + 2b_2 b_3 \sum X_{2i} X_{3i} \qquad (3.6)$$

نأخذ الشرط الأول *First order conditions* لتخفيض $\frac{\partial SSR}{\partial b_1} = 0$ و $\frac{\partial SSR}{\partial b_2} = 0$ و $\frac{\partial SSR}{\partial b_3} = 0$ ونحصل على المعادلات التالية:

$$\frac{\partial SSR}{\partial b_1} = -2 \sum_{i=1}^{n} (Y_i - b_1 - b_2 X_{2i} - b_3 X_{3i}) = 0 \qquad (3.7)$$

$$\frac{\partial SSR}{\partial b_2} = -2 \sum_{i=1}^{n} X_{2i} (Y_i - b_1 - b_2 X_{2i} - b_3 X_{3i}) = 0 \qquad (3.8)$$

$$\frac{\partial SSR}{\partial b_3} = -2 \sum_{i=1}^{n} X_{3i} (Y_i - b_1 - b_2 X_{2i} - b_3 X_{3i}) = 0 \qquad (3.9)$$

لدينا ثلاث معادلات بثلاث مجاهيل b_1 و b_2 و b_3، ونعيد ترتيب المعادلة الأولى لنحدد b_1 بناءً على b_2 و b_3 وبيانات Y و X_2 و X_3:
ويتم تحويل (3.7) كما يلي: (نقسم على 2)

$$\sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} b_1 + \sum_{i=1}^{n} b_2 X_{2i} + \sum_{i=1}^{n} b_3 X_{3i} \qquad (3.7a)$$

$$\sum_{i=1}^{n} Y_i = nb_1 + b_2 \sum_{i=1}^{n} X_{2i} + b_3 \sum_{i=1}^{n} X_{3i} \qquad (3.7b)$$

نقسم على n ونعرف $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$ نحصل على:

$$\bar{Y} = b_1 + b_2 \bar{X}_2 + b_3 \bar{X}_3 \qquad (3.7c)$$

ونحل المعادلة بالنسبة لـ b_1 ونحصل على:

$$\boldsymbol{b_1 = \bar{Y} - b_2 \bar{X}_2 - b_3 \bar{X}_3} \qquad \mathbf{(3.10)}$$

استخدم الصيغة (3.8) و (3.9) و (3.10) وحلها آنياً نحصل على صيغة b_2 التالية على شكل يتناسب مع صيغة انحراف المتغيّرات عن وسطها كما يلي: $x_{2i} = X_{2i} - \bar{X}_2$ و $x_{3i} = X_{3i} - \bar{X}_3$ و $y_i = Y_i - \bar{Y}$ تصبح الصيغة كما يلي:

$$b_2 = \frac{(\sum x_{2i} y_i)(\sum x_{3i}^2) - (\sum x_{3i} y_i)(\sum x_{2i} x_{3i})}{\sum x_{2i}^2 \sum x_{3i}^2 - (\sum x_{2i} x_{3i})^2} \qquad (3.11a)$$

ونحصل على صيغة b_3 بتبديل X_2 و X_3:

$$b_3 = \frac{(\sum x_{3i} y_i)(\sum x_{2i}^2) - (\sum x_{2i} y_i)(\sum x_{3i} x_{2i})}{\sum x_{2i}^2 \sum x_{3i}^2 - (\sum x_{2i} x_{3i})^2} \qquad (3.11b)$$

ونتوصل من هذا النقاش إلى نقطتين أساسيتين: الأولى: أن مبدأ اشتقاق معاملات الانحدار المتعدد هو نفسه للانحدار البسيط. والثانية: صيغة الحد الثابت b_1 هي توسيع لتحليل الانحدار البسيط، إلا أن صيغة معاملات الميل أكثر تعقيداً.

1.2.3- صيغة النموذج العام

عندما يكون لدينا اكثر من متغيّرين تفسيريين نستخدم جبر المصفوفات، ونفترض أن المتغيّر Y يعتمد على $k-1$ متغيّر تفسيري $X_2,\ldots,X_k$ حسب العلاقة الصحيحة غير المعلومة:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i \qquad (3.12)$$

لدينا مجموعة n مشاهدة لـ $Y, X_2,\ldots,X_k$ ونستخدم تحليل انحدار المربعات الصغرى لتقدير المعادلة:

$$\hat{Y}_i = b_1 + b_2 X_{2i} + \cdots + b_k X_{ki} \qquad (3.13)$$

وهذا يعني تخفيض مجموع مربعات البواقي المعطاة:

$$u_i = Y_i - \hat{Y}_i = Y_i - b_1 - b_2 X_{2i} - \cdots - b_k X_{ki} \qquad (3.14)$$

نختار الآن $b_1,\ldots,b_k$ لتخفيض مجموع مربعات البواقي (SSR) $\sum u_i^2$، ونحصل على k شرط أول $\partial SSR/\partial b_1 = 0$، ...،

$\partial SSR/\partial b_k = 0$ ونحصل على k معادلة لحل k مجهول، ونستطيع أن نرى بسهولة أولى هذه المعادلات الناتجة في حالة متغيّرين تفسيريين:

$$b_1 = \overline{Y} - b_2\overline{X}_2 - \cdots - b_k\overline{X}_k \qquad (3.15)$$

وبما أن صيغة $b_k,\ldots,b_2$ معقدة جداً، فإننا لا نستطيع عرضها رياضياً كما سبق لمتغيّرين، وينبغي إجراء تحليل بالجبر الخطي (المصفوفات).

$$Y_i = \beta_1 + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i \qquad (3.16)$$

بما أن $X_{1i} = 1$ لكل مشاهدة، سنستخدم الجبر العادي في التحليل، ونستطيع كتابة المعادلة (3.16) في صيغة المصفوفات كما يلي:

$$Y = X\beta + u \qquad (3.17)$$

حيث أن:

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \\ Y_T \end{pmatrix} \quad , \quad X = \begin{pmatrix} 1 & X_{21} & X_{31} & \cdots & X_{k1} \\ 1 & X_{22} & X_{32} & \cdots & X_{k2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & X_{2T} & X_{3T} & \cdots & X_{kT} \end{pmatrix}$$

$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \\ \beta_k \end{pmatrix} \quad , \quad u = \begin{pmatrix} u_1 \\ u_2 \\ \\ u_n \end{pmatrix}$$

لذا، فإن أبعاد المتجه Y أو مصفوفة $T\times 1$، و X مصفوفة $T\times k$، و β متجه $k\times 1$، و u متجه $T\times 1$، ويكون الهدف تخفيض SSR، و صيغة المصفوفة $SSR = u'u$، أي أن:

$$\begin{aligned} u'u &= (Y - X\hat{\beta})'(Y - X\hat{\beta}) \\ &= (Y' - \hat{\beta}'X')(Y - X\hat{\beta}) \\ &= Y'Y - Y'X\hat{\beta} - \hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta} \\ &= Y'Y - 2YX'\hat{\beta}' + \beta'X'X\hat{\beta} \end{aligned} \quad (3.18)$$

يتم اشتقاق الصيغة أعلاه بالنسبة لـ $\hat{\beta}$ ومساواتها بالصفر:

$$\frac{\partial SSR}{\partial \hat{\beta}} = -2X'Y + 2X'X\hat{\beta} = 0 \quad (3.19)$$

ونحصل على k معادلة، ونعيد كتابتها كما يلي:

$$X'X\hat{\beta} = X'Y \quad (3.20)$$

نضرب كلا الطرفين بمعكوس المصفوفة $(X'X)^{-1}$ ونحصل على:

$$\hat{\beta} = (X'X)^{-1}X'Y \quad (3.21)$$

وهي حل لمقدّرات المربعات الصغرى (OLS) في حالة تحليل الانحدار المتعدد.

مثال

لديك بيانات عن الكمية المنتجة من سلعة ما، وعوامل الإنتاج المستخدمة من عمل ورأس مال (القيم أرقام صحيحة لغايات تبسيط العمليات الحسابية في هذا المثال، في التطبيق العملي تؤخذ الصيغة اللوغارتمية لها) ولديك معادلة الإنتاج التالية:

$$q = \alpha + \beta_2 l + \beta_3 k \qquad (3.25)$$

المطلوب تقدير معلمات المعادلة اعلاه α و β_2 و β_3، وكذلك استخدم المصفوفات لاعادة تقديرها والتحقق من النتائج.

q	l	k
21	1	14
15	1	11
27	1	12
33	2	27
32	1	28

نجري العمليات الحسابية كما يلي:

Y	X_2	X_3				
q	l	k	$(Y_i - \bar{Y})$	$(X_{2i} - \bar{X}_2)$	$(X_{3i} - \bar{X}_3)$	$(X_{3i} - \bar{X}_3)^2$
21	1	14	-4.60	-0.2	-4.4	19.36
15	1	11	-10.60	-0.2	-7.4	54.76
27	1	12	1.40	-0.2	-6.4	40.96
33	2	27	7.40	0.8	8.6	73.96
32	1	28	6.40	-0.2	9.6	92.16
$\sum$=128	**6**	**92**	**0**	**0**	**0**	**281.2**
μ=25.6	**1.2**	**18.4**				

$_{2i}-\bar{X}_2)(Y_i-\bar{Y})$	$(X_{3i}-\bar{X}_3)(Y_i-\bar{Y})$	$(X_{2i}-\bar{X}_2)(X_{3i}-\bar{X}_3)$	$(X_{2i}-\bar{X}_2)^2$
0.92	20.24	0.88	0.04
2.12	78.44	1.48	0.04
-0.28	-8.96	1.28	0.04
5.92	63.64	6.88	0.64
-1.28	61.44	-1.92	0.04
7.4	**214.8**	**8.6**	**0.80**

$$b_2=\frac{(7.4\times 281.2)-(214.8\times 8.6)}{(0.8\times 281.2)-(8.6)^2}$$

$$=\frac{2080.88-1847.28}{224.96-73.96}=\frac{233.6}{151}\cong 1.547$$

$$b_3=\frac{(\sum x_{3i}y_i)(\sum x_{2i}^2)-(\sum x_{2i}y_i)(\sum x_{3i}x_{2i})}{\sum x_{2i}^2\sum x_{3i}^2-(\sum x_{2i}x_{3i})^2}$$

$$=\frac{(214.8)(0.8)-(7.4)(8.6)}{151}=\frac{(171.84)-(63.64)}{151}$$

$$=\frac{108.2}{151}=0.716556\cong 0.717$$

$$b_1=\bar{Y}-b_2\bar{X}_2-b_3\bar{X}_3$$
$$=25.6-(1.547)(1.2)-(0.717)(18.4)$$
$$=10.55$$

وبناءً علية تكون المعادلة كما يلي: $\hat{Q}=10.55+1.547\,L+0.717\,K$

ثانياً ولإعادة تقديرها بطريقة المصفوفات نستخدم الصيغة التالية

$$\beta_i=(X'X)^{-1}\,X'Y$$

نتبع الإجراءات التالية:

$$(X'X)=\begin{bmatrix}1&1&1&1&1\\1&1&1&2&1\\14&11&12&27&28\end{bmatrix}\times\begin{bmatrix}1&1&14\\1&1&11\\1&1&12\\1&2&27\\1&1&28\end{bmatrix}$$

$$=\begin{pmatrix}5&6&92\\6&8&119\\92&119&1974\end{pmatrix}$$

$$(X'Y)=\begin{bmatrix}1&1&1&1&1\\1&1&1&2&1\\14&11&12&27&28\end{bmatrix}\times\begin{bmatrix}21\\15\\27\\33\\32\end{bmatrix}=\begin{bmatrix}128\\161\\2570\end{bmatrix}$$

$$(X'X)^{-1}=\begin{pmatrix}\frac{1631}{755}&\frac{-896}{755}&\frac{-22}{755}\\\frac{-896}{755}&\frac{1406}{755}&\frac{-43}{755}\\\frac{-22}{755}&\frac{-43}{755}&\frac{4}{755}\end{pmatrix}$$

$$\beta_i=(X'X)^{-1}\;X'Y=\begin{pmatrix}\beta_1\\\beta_2\\\beta_3\end{pmatrix}=\begin{pmatrix}\frac{7972}{755}\\\frac{1168}{755}\\\frac{541}{755}\end{pmatrix}=\begin{pmatrix}10.5589\\1.5470\\0.7165\end{pmatrix}$$

وهي مطابقة للنتائج أعلاه، مع اختلاف بسيط يعود لعوامل التقريب.

ثالثاً: ولمزيد من التأكد من النتائج تم حساب المعادلة من خلال برنامج EViews 9.0 وكانت النتائج كما يلي:

Dependent Variable: Q
Method: Least Squares
Date: 11/24/15 Time: 22:13
Sample: 2001 2005
Included observations: 5

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	10.55894	8.432754	1.252134	0.3371
L	1.547020	7.829520	0.197588	0.8616
K	0.716556	0.417612	1.715844	0.2283

R-squared	0.715243	Mean dependent var	25.60000
Adjusted R-squared	0.430486	S.D. dependent var	7.602631
S.E. of regression	5.737411	Akaike info criterion	6.615602
Sum squared resid	65.83576	Schwarz criterion	6.381265
Log likelihood	-13.53901	Hannan-Quinn criter.	5.986664
F-statistic	2.511769	Durbin-Watson stat	2.761440
Prob(F-statistic)	0.284757		

وهي مطابقة للناتج أعلاه تماماً.

3-3- خصائص معاملات الانحدار المتعدد

تعتبر معاملات الانحدار في تحليل الانحدار البسيط حالة خاصة لمتغيّرات عشوائية مكوّنها العشوائي يعزى لوجود حد الخطأ في النموذج. وكل معامل انحدار يحسب كدالة في قيم Y والمتغيّرات التفسيرية، ويتحدد Y بالمتغيّرات التفسيرية وحد الخطأ، وتتبع معاملات الانحدار التي تتحدد

بقيم المتغيّرات التفسيرية وحد الخطأ، وتعتمد خصائصها على خصائص الأخيرة.

3-3-1- فرضيات نموذج الانحدار المتعدد

سنعمل في إطار نموذج متغيّراته التفسيرية غير عشوائية، وسنعيد صياغة الفرضيات كما في الفصل الثاني لتناسب نموذج الانحدار المتعدد.

1- المتغيّر التابع دالة خطية في المتغيّرات التفسيرية والدالة صحيحة الوصف

$$Y = \beta_1 + \beta_2 X_2 + \cdots + \beta_k X_k + u$$

كما في السابق باستثناء احتوائه متغيّرات تفسيرية متعددة.

2- جميع المتغيّرات التفسيرية غير عشوائية

3- جميع المتغيّرات التفسيرية لها قيم ثابتة عند تكرار العينة.

4- توقع حد الخطأ صفر:

$$E(u_i) = 0, \qquad \forall i$$

5- تجانس تباين حد الخطأ Homoskedastic

$$\sigma_{ui}^2 = \sigma_u^2 = \text{ثابت}$$

6- استقلال قيم حد الخطأ: توزيع u_i مستقل عن u_j لكل قيم $i \neq j$

$$Cov(u_i, u_j) = E(u_i, u_j) = 0, \quad \forall i \neq j$$

7- توزيع حد الخطأ طبيعي: جميع قيم u_i تتوزع طبيعياً.

8- عدم وجود علاقة خطية دقيقة بين متغيّرين تفسيريين أو أكثر (لا يوجد ارتباط خطي متعدد Multicollinearity).

3-3-2- مصفوفة التباين- والتباين المشترك للأخطاء The Variance-Covariance matrix of the errors

يعطينا تحليل التباين-التباين المشترك لمعلمات المربعات الصغرى معلومات حول دقة reliability المقدّرات b_1 و b_2 و b_3. وبما أن معلمات المربعات الصغرى غير متحيّزة؛ فإن صغر تباينها يزيد من احتمالية إنتاج تقديرات قريبة من قيم المعلمات الصحيحة، وزيادة تباين الخطأ σ^2 يؤدي إلى زيادة تباين معلمات المربعات الصغرى وهو المتوقع، حيث يقيس σ^2 حالة عدم التأكد في توصيف النموذج. وإذا كان التباين σ^2 كبيراً، سيكون انتشار قيم البيانات واسعاً لدالة الانحدار $E(Y_i) = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3}$ ونحصل على معلومات أقل من بيانات قيم المعلمات، وإذا كان σ^2 أقل من قيم البيانات يكون الانتشار مضغوطاً حول دالة الانحدار $E(Y_i) = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3}$ ونحصل على معلومات أكثر عما ستكون عليه قيم المعلمات.

ويتم تنظيم التباين-التباين المشترك لمعلمات المربعات الصغرى على شكل مجموعة مربعة تسمى مصفوفة Matrix، ويكون التباين في قطر هذه المصفوفة، والتباين المشترك حول القطر؛ وتسمى هذه المصفوفة بمصفوفة **التباين-التباين المشترك Variance-Covariance matrix** أو **مصفوفة التباين المشترك Covariance matrix**، وعندما تكون $k=3$ يكون التباين والتباين المشترك في مصفوفة التباين كما يلي:

$$\operatorname{cov}(b_1, b_2, b_3) = \begin{bmatrix} \operatorname{var}(b_1) & \operatorname{cov}(b_1, b_2) & \operatorname{cov}(b_1, b_3) \\ \operatorname{cov}(b_1, b_2) & \operatorname{var}(b_2) & \operatorname{cov}(b_2, b_3) \\ \operatorname{cov}(b_1, b_3) & \operatorname{cov}(b_2, b_3) & \operatorname{var}(b_3) \end{bmatrix} \quad (3.22)$$

باستخدام برمجية EViews يكون التباين والتباين المشترك المقدّر للمعلمات b_1 و b_2 و b_3 في مثال الطلب على النقود الضيق $(M1)$ كما يلي:

$$\hat{\text{cov}}(b_1, b_2, b_3) = \begin{bmatrix} 0.068386 & -0.006280 & -0.009649 \\ -0.006280 & 0.000665 & 0.000620 \\ -0.009649 & 0.000620 & 0.002172 \end{bmatrix}$$

لذا يكون لدينا:

$$\text{var}(b_1) = 0.068386 \qquad \text{cov}(b_1, b_2) = -0.006280$$

$$\text{var}(b_2) = 0.000665 \qquad \text{cov}(b_1, b_3) = -0.009649$$

$$\text{var}(b_3) = 0.002172 \qquad \text{cov}(b_2, b_3) = 0.000620$$

جدول (3-2) تقدير معاملات مصفوفة التباين

	C	LOG(GDP)	LOG(R)
C	0.068386	-0.006280	-0.009649
LOG(GDP)	-0.006280	0.000665	0.000620
LOG(R)	-0.009649	0.000620	0.002172

يظهر الجدول (3-2) المعلومات في تقرير نمطي لنتائج الحاسوب. ونأخذ الجذر التربيعي للتباينات المقدّرة ونحصل على الخطأ المعياري للمعلمات b_1 و b_2 و b_3 كما يلي:

$$se(b_1) = \sqrt{\text{var}(b_1)} = \sqrt{0.068386} = 0.2615$$

$$se(b_2) = \sqrt{\text{var}(b_2)} = \sqrt{0.000665} = 0.0257$$

$$se(b_3) = \sqrt{\text{var}(b_3)} = \sqrt{0.002172} = 0.0466$$

أنظر إلى الجدول (3-2) ولاحظ أن تلك القيم تظهر في عمود الخطأ المعياري في نتائج الحاسوب.

تمرين (1) لدينك البيانات التالية

قدّر معاملات الانحدار المتعدد واستخدم صيغة انحراف المتغيّرات عن وسطها لغايات التبسيط: $x_{2i} = X_{2i} - \overline{X}_2$ و $x_{3i} = X_{3i} - \overline{X}_3$ و $y_i = Y_i - \overline{Y}$

year	Y	X_2	X_3	y	x2	x3	x2y	x3y	x2x3	x_2^2	x_3^2
2000	40	6	4								
2001	44	10	4								
2002	46	12	5								
2003	48	14	7								
2004	52	16	9								
2005	58	18	12								
2006	60	22	14								
2007	68	24	20								
2008	74	26	21								
2009	80	32	24								
N=10	$\sum$=570	$\sum$=180	$\sum$=120								
	μ=57	μ=18	μ=12								

تكون النتيجة $\hat{Y}_i = 31.98 + 0.65 X_{1i} + 1.11 X_{2i}$

تمرين (2)

احسب قيم t لمعاملات الانحدار المتعدد وفسر وعلق على المعادلة التالية:

Dependent Variable: Y
Method: Least Squares
Date: 11/25/15 Time: 21:48
Sample: 2000 2009
Included observations: 10

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	31.98067	1.631796		0.0000
X2	0.650051	0.250161		0.0355
X3	1.109868	0.267434		0.0043

R-squared	0.991634	Mean dependent var	57.00000
Adjusted R-squared	0.989243	S.D. dependent var	13.47426
S.E. of regression	1.397467	Akaike info criterion	3.750525
Sum squared resid	13.67040	Schwarz criterion	3.841300
Log likelihood	-15.75262	Hannan-Quinn criter.	3.650944
F-statistic	414.8492	Durbin-Watson stat	2.114085
Prob(F-statistic)	0.000000		

3-4- خصائص مقدّرات نموذج المربعات الصغرى للانحدار المتعدد

كما في نموذج الانحدار بمتغيّرين نستطيع اثبات أن مقدّرات OLS هي أفضل مقدّرات خطية غير منحازة Best Linear Unbiased Estimators (BLUE) مركزين على اثبات أن معاملات الميل $(\beta_k,\ldots,\beta_3,\beta_2)$ بدلاً من الثابت β_1 لأن تلك المعاملات لها أهمية أكبر.

1- الخطية Linearity

مقدّرات OLS خطية، وبما أن قيم المتغيّرات التفسيرية مقطعها ثابت نستطيع أن نرى أن مقدّرات OLS دالة خطية بقيم Y، ويكون حل $\hat{\beta}$ كما يلي:

$$\hat{\beta}=(X'X)^{-1}X'Y \qquad (3.23)$$

وبما أن المصفوفة حدها الثابت ثابت، فإن $\omega=(X'X)^{-1}X'$ كذلك المصفوفة $n\times k$ مقطعها ثابت *Fixed constants*، وبما أن ω مصفوفة الحد ثابت؛ فإن $\hat{\beta}$ دالة خطية في Y، وبالتالي مقدّراتها خطية.

2- عدم التحيّز Unbiasedness

لأن $E(\beta)=\beta$ و $E(u)=0$، فإن $\hat{\beta}$ مقدّر غير منحاز لـ β.

وهذا ما نثبته كما يلي:

$$E\left(\hat{\beta}_2\right)=E\left[\beta_2+\frac{\sum\left(X_t-\bar{X}\right)u_t}{\sum\left(X_t-\bar{X}\right)^2}\right]$$

$$=\beta_2+E\left[\frac{\sum\left(X_t-\bar{X}\right)u_t}{\sum\left(X_t-\bar{X}\right)^2}\right] \qquad (3.24)$$

فإذا كان المتغيّر التفسيري X_t مستقلاً عن u_t فإن:

$$E\left[\frac{\sum\left(X_t-\bar{X}\right)u_t}{\sum\left(X_t-\bar{X}\right)^2}\right]=\frac{E\left[\sum\left(X_t-\bar{X}\right)\right]E\left[u_t\right]}{E\left[\sum\left(X_t-\bar{X}\right)^2\right]} \qquad (3.25)$$

وبما أن $E(u_t)=0$ فإن:

$$E(\hat{\beta}_2)=\beta_2+0$$
$$=\beta_2 \quad (3.26)$$

وبالتالي فإن تقدير المربعات الصغرى سيكون تقديراً غير متحيّز على فرض أن المتغيّر التفسيري X_t مستقلاً عن حد الخطأ u_t.

3- الاتساق Consistency

الاتساق يعني ببساطة أن تقدير $\hat{\beta}$ سيساوي β الصحيحة، وهذا يعني أن التقدير يذهب إلى ما لا نهاية وستتقارب $\hat{\beta}$ من القيمة الصحيحة $Plim(\hat{\beta})=\beta$.

3-5- جودة التقدير

3-5-1- معامل التحديد R^2 و R^2 المصحح

كما في الانحدار البسيط، يقيس معامل التحديد R^2 نسبة فروقات المتغيّر التابع Y المُفَسَّرة بفروقات المتغيّر التفسيري، ونقدم في الانحدار المتعدد نفس المقياس ونفس الصيغة، لكننا سنتكلم عن نسبة فروقات المتغيّر التابع المُفَسَّرة بجميع المتغيّرات التفسيرية الداخلة بالنموذج الخطي، وبذلك يكون معامل التحديد Coefficient of determination كما يلي:

$$R^2 = \frac{SSE}{SST} = \frac{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} \quad (3.27)$$

أو

$$R^2 = 1 - \frac{SSR}{SST} = 1 - \frac{\sum_{i=1}^{n} u_i^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} \quad (3.28)$$

حيث أن SSE هي فروقات Y "المُفَسَّرة" في النموذج (مجموع مربعات الانحدار)، و SST هي مجموع فروقات Y حول وسطها (مجموع المربعات الكلي)، و SSR هي مجموع المربعات الصغرى (الأخطاء أو البواقي risdual) وهي فروقات Y غير المُفَسَّرة في النموذج.

تشير $\hat{Y}$ إلى القيمة المتنبأ بها predicted value للمتغيّر التابع Y لكل قيم المتغيّرات التفسيرية، حيث أن:

$$\hat{Y}_i = b_1 + b_2 X_{2i} + b_3 X_{3i} + \cdots + b_k X_{ki} \quad (3.29)$$

الوسط الحسابي للعينة $\bar{Y}$ هو وسط كل من Y_i و $\hat{Y}_i$، يزودنا بنموذج يتضمن المقطع b_1 في هذه الحالة.

تعطينا جميع برمجيات الحاسوب قيمة SSR، إلا أنه في بعض الأحيان لا تظهر SST، وعليه يُعطى الخطأ المعياري للعينة وهو محسوب في أغلب البرمجيات كما يلي:

$$\hat{\sigma}_Y = \sqrt{\frac{1}{N-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2} = \sqrt{\frac{SST}{N-1}} \quad (3.30)$$

وبالتالي، فإن:

$$SST = (N-1)\hat{\sigma}_Y^2 \qquad (3.31)$$

وفي مثال الطلب على النقود، نجد أن $SST = 53 \times (0.040673)^2 = 2.155669$، و $SSR = 0.084368$،

وباستخدام مجموع المربعات يكون لدينا:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} u_i^2}{\sum_{i=1}^{n} (Y_i - \overline{Y})^2} = 1 - \frac{0.084368}{2.155669} = 0.96 \qquad (3.32)$$

يُعبَّر عن R^2 بأن 96٪ من فروقات الطلب على النقود يفسر بفروقات الدخل وفروقات سعر الفائدة، وتعني أن 4٪ من فروقات الطلب على النقود بقيت غير مُفَسَّرة نتيجة فروقات حد الخطأ أو فروقات المتغيّرات الأخرى التي هي ضمناً جزءاً من حد الخطأ.

يعرض معامل التحديد كذلك مقياساً لقدّرة النموذج على التنبؤ خلال فترة العينة، أو قياس جودة الانحدار المقدّر، وتساوي قيمة R^2 مربع معامل ارتباط العينة بين Y_i و $\hat{Y}_i$، حيث أن الارتباط يقيس العلاقة الخطية المناسبة بين المتغيّرين. فإذا كان R^2 مرتفعاً، فهذا يعني أن الترابط بين القيمة Y_i والقيمة المتنبأ بها $\hat{Y}_i$ يكون قريباً، وفي هذه الحالة يكون نموذج تقدير البيانات جيداً، فإذا كان R^2 منخفضاً لا يكون الترابط قريباً بين قيم Y_i وقيم $\hat{Y}_i$ المتنبأ بها في النموذج، ولا يكون النموذج مناسباً.

احدى صعوبات استخدام R^2 كمقياسٍ لجودة التقدير هي زيادة حجمة عند اضافة متغيّرات جديدة حتى لو كانت المتغيّرات المضافة ليس

لها أي قيمة اقتصادية، وجبرياً فإن اضافة المتغيّرات تخفض من مجموع مربعات الأخطاء SSR وبالتالي يرتفع R^2، فإذا احتوى النموذج $N-1$ متغيّراً يكون $R^2 = 1$، وليس من الحكمة الحصول على قيمة R^2 مرتفعة.

وكمقياس بديل لجودة التقدير هو R^2 المصحح adjusted ويرمز له $\overline{R}^2$ ويظهر عادة في برامج الانحدار ويحسب كما يلي:

$$\overline{R}^2 = 1 - \frac{SSR/(N-K)}{SST/(N-1)} \tag{3.33}$$

تُظهر بيانات الطلب على النقود أن $\overline{R}^2 = 0.965$، وهذا المقياس لا يزيد دائماً عندما تضاف متغيّرات جديدة بسبب وجود درجات الحرية $N-K$ في البسط، مثلاً عند زيادة عدد المتغيّرات K ينخفض SSR وبالتالي $N-K$ تنخفض، ويعتمد أثر $\overline{R}^2$ على حجم انخفاض SSR. ونحاول تعويض هذا التحول التصاعدي التلقائي بفرض عقوبة عن زيادة عدد المتغيّرات التفسيرية وتحدد كما يلي:

$$\begin{aligned}\overline{R}^2 &= 1-\left(1-R^2\right)=\frac{n-1}{n-k}R^2-\frac{k-1}{n-k}\\ &= R^2-\frac{k-1}{n-k}\left(1-R^2\right)\end{aligned} \tag{3.34}$$

حيث $k-1$ عدد المتغيّرات التفسيرية، وبما أن k تزداد، فإن $\frac{k-1}{n-k}$ تزداد كذلك، وبالتالي يزداد التصحيح السلبي لـ R^2.

3-5-2- اختبار معنوية المعلمات الفردية

يزودنا توزيع t بأساس لاختبار فرضيات المعلمات الفردية، وكما هو في الفصل (2) تأخذ الفرضية الشكل $H_0: \beta_2 = c$ مقابل $H_1: \beta_2 \neq c$؛ حيث أن c هي ثابت، وتسمى هذه الفرضية باختبار بذيلين two-tail test. أما فرضية عدم المساواة التالية: $H_0: \beta_2 \leq c$ مقابل $H_1: \beta_2 > c$ تسمى اختبار بذيل واحد one-tail test، وسنأخذ بعض الأمثلة لكل نوع من هذه الفرضيات: فقد نستخدم اختبار بذيلين لاختبار معنوية المعلمات الفردية، ويستخدم اختبار بذيل واحد لاختبار بعض الفرضيات الاقتصادية. ولاختبار الفرضيات سنتبع الإجراء خطوة بخطوة، ولتنشيط ذاكرتنا نكرر نفس خطوات الاختبار التالية:

1- حدد الفرضية الأساسية والفرضية البديلة.

2- حدد الاختبار الإحصائي وتوزيعه إذا كانت الفرضية الأساسية صحيحة.

3- اختر مستوى المعنوية α وحدد منطقة الرفض.

3- احسب احصائية t للعينة والقيمة الاحتمالية.

5- حدد النتيجة.

نعتقد أن المتغيّرات المستقلة تؤثر في المتغير التابع Y في نموذج الانحدار المتعدد، وإذا أردنا التأكد من هذا الاعتقاد سنحتاج أن نختبر فيما إذا كانت البيانات المستخدمة في التحليل تدعمه أم لا، ونتساءل فيما إذا

كانت البيانات تزودنا بأي أدلة تبين العلاقة بين Y وكل متغيّر تفسيري، فإذا كان المتغيّر التفسيري مثل X_K لا يؤثر على Y سنستنتج أن $\beta_K = 0$ ، ويسمى اختبار الفرضية الأساسية باختبار الدلالة (أو المعنوية) للمتغيّر التفسيري X_K ، ثم الكشف عن أدلة تدعم وجود علاقة تربط Y مع X_K ، ونختبر الفرضية الأساسية التالية:

$$H_0 : \beta_K = c$$

مقابل الفرضية البديلة:

$$H_1 : \beta_K \neq c$$

نستخدم إحصائية t لاختبار صحة الفرضية الأساسية كما يلي:

$$t = \frac{b_K}{se(b_K)} \sim t_{(N-K)} \qquad (3.35)$$

وفي الفرضية البديلة "عدم المساواة" نستخدم اختبار بذيلين، حيث نرفض H_0 إذا كانت قيمة t المحسوبة أكبر من/ أو تساوي $t_{\text{الحرجة}}$ (القيمة الحرجة من الجانب الأيمن للتوزيع) أو أقل من/ أو تساوي $-t_{\text{الحرجة}}$ (القيمة الحرجة من الجانب الأيسر للتوزيع) عند مستوى معنوية α، أي $t_{\text{الحرجة}} = t_{(1-\alpha/2, N-K)}$ و $-t_{\text{الحرجة}} = t_{(\alpha/2, N-K)}$ أو استخدام صيغة قاعدة القبول- الرفض حسب القيمة الاحتمالية p-value حيث سنرفض H_0 إذا كانت $p \leq \alpha$ ، ولا نرفض H_0 إذا $p > \alpha$.

وفي مثال دالة الطلب على النقود سنختبر فيما إذا كان الطلب على النقود يعتمد على الناتج المحلي الإجمالي كما يلي:

1- الفرضية الأساسية والفرضية البديلة هما: $H_0 : \beta_2 = 0$ و $H_1 : \beta_2 \neq 0$.

2- استخدام احصائية اختبار صحة الفرضية الأساسية: $t = b_2 / se(b_2) \sim t_{(N-K)}$.

3- استخدم مستوى معنوية 5٪ $(\alpha = 0.05)$، مع ملاحظة أن درجات الحرية 51 ستكون القيمة الحرجة التي تحقق احتمالية 0.025 لكل ذيل من التوزيع هي: $t_{(0.975,51)} = 2.000$ و $t_{(0.025,51)} = -2.000$، لذا سنرفض الفرضية الأساسية إذا كانت قيمة t المحسوبة من الخطوة (2) تمثل $t \geq 2.000$ أو $t \leq -2.000$، أما إذا كانت $-2.000 < t < 2.000$ لا نستطيع رفض H_0، أما قاعدة القبول/ الرفض حسب مفهوم القيمة الاحتمالية: نرفض H_0 إذا كانت $p \leq 0.05$ ولا نستطيع رفض H_0 إذا كانت $p > 0.05$.

3- قيمة احصائية t المحسوبة هي:

$$t = \frac{0.362177}{0.025780} = 14.04867$$

تُظهر نتائج الحاسوب أن القيمة الاحتمالية التي نستطيع ايجادها كما يلي:

$$P\left(t_{(51)} > 14.04867\right) + P\left(t_{(51)} < -14.04867\right) = 0.0000$$

وتكون القيمة الاحتمالية= 0.0000

5- بما أن $14.04867 > 2.000$ سنرفض $H_0 : \beta_2 = 0$ ونستنتج أن البيانات تعطي دليلاً على أن الطلب على النقود يعتمد على

الدخل. وباستخدام القيمة الاحتمالية لتنفيذ الاختبار سنرفض H_0 لأن $0.0000 < 0.05$.

ولاختبار فيما إذا كان الطلب على النقود يعتمد على السعر نتبع نفس الخطوات اعلاه. ونترك هذا الاختبار كتمرين.

تطبيق

أ- اختبار فاعلية الدخل

سيتم اختبار فرضية أن الدخل يزيد من الطلب على النقود، وحيث أن الزيادة ستحقق $\beta_2 > 1$ سيتم تشكيل الفرضية كما يلي:

1- $H_0 : \beta_2 \leq 1$ و $H_1 : \beta_2 > 1$.

2- عالج الفرضية الأساسية بالمساواة $H_0 : \beta_2 = 1$، وتكون إحصائية الاختبار التي توزيعها t عند H_0 كما يلي:

$$t = \frac{b_2 - 1}{se(b_2)} \sim t_{(N-K)} \qquad (3.36)$$

3- اختر $\alpha = 0.05$ كمستوى معنوية، وبذلك تكون القيمة الحرجة لها هي $t_{(0.95,51)} = 1.676$، وسنرفض H_0 إذا كانت $t \geq 1.676$ أو القيمة الاحتمالية $p-value \leq 0.05$.

4- قيمة احصائية الاختبار هي:

$$t = \frac{b_2 - \beta_2}{se(b_2)} \qquad (3.37)$$

$$t = \frac{0.362177 - 1}{0.025780} = -24.741$$

والقيمة الحرجة للاختبار $P(t_{(51)} > 1.263) = 0.005$.

5- بما أن $-24.741 < -1.676$ سنرفض H_0 ونحصل على دليل بأن الدخل سيكون فعّالاً في التأثير على الطلب على النقود، وباستخدام القيمة الاحتمالية نستنتج مرة أخرى أن H_0 لا نستطيع قبولها لأن $0.005 < 0.05$ ، وقد يتسآل سائل لماذا نقر بهذه النتيجة بالرغم من أن قيمة β_2 هي أقل من الواحد، والجواب هو أن هذه القيمة هي قيمة المرونة التي تشير إلى أن معلمة الدخل غير مرنة (لأن البيانات بصيغة اللوغاريتم)، وإذا أرجعنا قيمة المرونة إلى القيمة الأصلية لها كما يلي:

$$\hat{\varepsilon} = b_2 \times \frac{\overline{X}}{\overline{Y}} \tag{3.38}$$

$$0.362177 = b_2 \times \frac{1422.254}{6058.839} = b_2 \times 0.2347404$$
$$\Rightarrow b_2 = 1.543$$

نستنتج أن $\beta_2 > 1$.

ب- اختبار مرونة الطلب على النقود

نرغب بمعرفة ما يتعلق بمرونة الطلب:

- $\beta_3 \geq 0$: انخفاض سعر الفائدة يؤدي إلى انخفاض الطلب على النقود (الطلب غير مرن السعر)، أو

- $\beta_3 < 0$: انخفاض سعر الفائدة يؤدي إلى زيادة الطلب على النقود (الطلب مرن السعر).

إذا لم نقبل بفرضية أن الطلب على النقود مرن، سيكون لدينا دليلاً قوياً على أن البيانات تدعم هذا الإدعاء، وسيكون من المناسب أخذ فرضية عدم مرونة الطلب كفرضية أساسية، وفيما يلي نشكل الاختبار المعياري، وسنعرض أولاً الفرضية الأساسية والبديلة:

1- $H_0 : \beta_3 \geq 0$ (الطلب مرونته واحد أو غير مرن).

$H_1 : \beta_3 < 0$ (الطلب مرن).

2- لتكوين احصائية الاختبار سنفترض أن الفرضية الأساسية $\beta_3 = 0$ من الاحصائية $t = b_3/se(b_3) \sim t_{(N-K)}$.

3- تتوافق منطقة الرفض مع قيمة توزيع t (القيمة الحرجة $t_{(0.05,51)} = -1.676$) وسنرفض H_0 إذا كانت $t \leq -1.676$ أو $p-value < 0.05$.

4- قيمة إحصائية الاختبار:

$$t = \frac{b_3}{se(b_3)} = \frac{-1.092}{0.0466} = -23.44$$

وبالمثل القيمة الاحتمالية هي $P(t_{(51)} \leq -23.44) = 0.0000$.

5- بما أن $-23.44 < -1.676$ سنرفض $H_0 : \beta_3 \geq 0$، ونستنتج أن $H_1 : \beta_3 < 0$ (الطلب مرن)، وتدعم الأدلة أن انخفاض سعر

الفائدة سيجلب زيادة الطلب على النقود، وبما أن $0.05 > 0.0000$ نفسها النتيجة التي توصلنا إليها باستخدام القيمة الاحتمالية.

3-5-3- **فترات التقدير**

افرض أننا مهتمين في ايجاد فترة تقدير باحتمال 95٪ للمعلمة β_2 لاستجابة الطلب على النقود للناتج المحلي الاجمالي، وباتباع الاجراءات الموصوفة في (2-9-3) وملاحظة وجود $N-K=54-3=51$ درجة حرية، وتكون الخطوة الأولى ايجاد القيمة الحرجة $t_{\text{الحرجة}}$ لتوزيع t_{51}:

$$P(-t_{\text{الحرجة}} < t_{51} < t_{\text{الحرجة}}) = 0.95 \qquad (3.39)$$

القيمة الحرجة $t_{\text{الحرجة}} = t_{(0.975,\, N-K)}$ هي باحتمال 97.5 للتوزيع $t_{(N-K)}$ (المساحة أو الاحتمال على يسار $t_{\text{الحرجة}}$ هو 0.975) و $-t_{\text{الحرجة}} = t_{(0.025,\, N-K)}$ هو احتمال 2.5 للتوزيع $t_{(N-K)}$ (المساحة أو الاحتمال على يسار $-t_{\text{الحرجة}}$ هو 0.025)، وبالنظر في جدول t سنكتشف عدم وجود درجات حرية 51، لكنها تقع بين درجات حرية 40 و 60، ومن الواضح أنها مصححة بثلاث خانات (بثلاث منازل عشرية) $t_{\text{الحرجة}} = 2.000$، استخدم (3.37) للمعلمة الثانية في المعادلة $(k=2)$ نستطيع اعادة كتابة (3.39) كما يلي:

$$P\left(-2.000 \leq \frac{b_2 - \beta_2}{se\,(b_2)} < 2.000\right) = 0.95 \qquad (3.40)$$

أعد ترتيب (3.40) نحصل على:

$$P[b_2 - 2.000 \times se(b_2) \leq \beta_2 < b_2 + 2.000 \times se(b_2)] = 0.95$$

وتكون الفترة:

$$[b_2 - 2.000 \times se(b_2) \quad , \quad b_2 + 2.000 \times se(b_2)] \qquad (3.41)$$

حدد احتمال 95٪ للفترة تقدير المعلمة β_2، وإذا تم استخدام فترة التقدير في عدة عينات من المجتمع، سيحتوي احتمال 95٪ على معلمة β_2 الصحيحة، ونستطيع انجاز هذه الحقيقة قبل جمع أي بيانات بالاعتماد على نموذج الفرضيات وحده، وقبل جمع البيانات لدينا ثقة في فترة إجراءات التقدير.

نحصل من (3.41) على فترة تقدير المعلمة β_2 باحتمال 95٪ بتعويض b_2 و $se(b_2)$ بالقيمة $b_2 = 0.3622$ و $se(b_2) = 0.0\ 2\ 5$، ونحصل على فترة تقدير للمعلمة β_2 باحتمال 95٪ التالية:

$$[0.3622 - 2.000 \times 0.0258 \quad , \quad 0.3622 + 2.000 \times 0.0258]$$

$$[0.3106 \quad , \quad 0.4138]$$

تبين فترة التقدير أن زيادة الناتج المحلي الاجمالي بمقدار 1 مليون دينار سيؤدي إلى زيادة الطلب على النقود بين 0.3106 مليون دينار و 0.4138 مليون دينار، أو بمفهوم التغيّر في الناتج المحلي الاجمالي الذي يعني أن زيادة الناتج المحلي الاجمالي بنسبة 10٪ تزيد الطلب على النقود بين 3.11 - 4.14 ألف دينار، وبالاعتماد على هذه المعلومات فإن كمية الطلب على النقود تعتمد على زيادة الدخل.

ونتبع نفس الإجراءات للمعلمة β_3 لاستجابة الطلب على النقود على سعر الفائدة، نجد فترة تقدير باحتمال 95٪:

$$[-1.0924-2.000\times 0.0466, -1.0924+2.000\times 0.0466]$$
$$=[-1.186, \quad -0.999]$$

نقدّر الزيادة في سعر الفائدة 10٪ تؤدي إلى انخفاض الطلب على النقود بين 9.99٪ و 11.86٪، وهي تعني أن زيادة سعر الفائدة قد يخفض الطلب (الطلب ينخفض بأقل من 0.0999 دينار) أو قد يؤدي إلى انخفاض الطلب بأقل من الفائدة 11.86٪، وطريقة أخرى لوصف الحالة نقول أن نقطة تقدير $b_3 = 1.0924$ موثوقه لأن انحرافها المعياري صغير نسبياً.

للحصول على صيغة عامة لفترة تقدير نحتاج معرفة القيمة $t_{\text{الحرجة}}$ اعتماداً على درجة الثقة المحددة لفترة التقدير وعدد درجات الحرية، ونشير إلى درجة الثقة $1-\alpha$ في حالة احتمال الفترة المقدّرة باحتمال 95٪؛ تكون $\alpha = 0.05$ و $1-\alpha = 0.95$، ودرجة الحرية هي $N-K$ التي كانت في مثال الطلب على النقود $54-3=51$ والقيمة الحرجة ($t_{\text{الحرجة}}$) هي احتمال $t_{(1-\alpha/2, N-K)}$، التي هي $P\left[t_{(N-K)} \leq t_{(1-\alpha/2, N-K)}\right] = 1-\alpha/2$ ، وفي حالة فترة ثقة باحتمال 95٪، $1-\alpha/2 = 0.975$، نستخدم هذه القيمة لأنها يتطلب 0.025 في كل ذيل للتوزيع، وبالتالي نكتب الصيغة العامة لفترة الثقة باحتمال $100(1-\alpha)\%$ كما يلي:

$$\left[b_k - t_{(1-\alpha/2, N-K)} \times se(b_k), \quad b_k + t_{(1-\alpha/2, N-K)} \times se(b_k)\right] \qquad (3.42)$$

3-5-4- اختبار F

تعلمنا كيفية استخدام اختبار t لاختبار فرضيات المعلمات الفردية في نموذج الانحدار المتعدد. أما إذا اردنا اختبار أكثر من معلمة بنفس الوقت ينبغي تضمين النموذج العديد من المتغيّرات كمجموعة متغيّرات تفسيرية مثل نموذج كمية الطلب، فيما إذا كان يعتمد على أسعار السلع البديلة أم على أسعارها فقط، وهذا يقودنا إلى التساؤل عن اختبار فرضية تتضمن أكثر من معلمة؛ إلا أنه لا يتضمن اختبار مجموعة متغيّرات، وهل تبين دالة الإنتاج ثبات عوائد الحجم؟ فإذا ارتفعت جميع الأسعار والدخل بنفس النسبة، فهل تبقى الكمية المطلوبة من السلعة ثابتة؟

سيتم التمييز بين الفرضية الأساسية الفردية Single null hypothesis حيث تكون الفرضية الأساسية مقيّدة بقيد واحد على معلمة واحدة أو أكثر، والفرضية الأساسية المشتركة Joint null hypothesis التي تتضمن قيدين أو أكثر على معلمتين أو اكثر؛ حيث يتم تطبيق اختبار الفرضية الأساسية الفردية بذيلين من خلال اختبار t أو اختبار F وهما متكافئان، واختبار الفرضية الأساسية بذيل واحد يجب استخدام اختبار t، ويجب استخدام اختبار F لاختبار الفرضية الأساسية المشتركة.

إذا أردنا استخدام اختبار F لاختبار قوة المتغيّرات التفسيرية المشتركة في نموذج الانحدار المتعدد، نأمل برفض الفرضية الأساسية التي تقول بأن النموذج لا تحوي متغيّراته قوة تفسيرية عندما لا يكون للمتغيّر Y علاقة له بأي متغير تفسيري، ويعبّر عنه رياضياً كما يلي:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i \tag{3.43}$$

وتكون الفرضية الأساسية H_0 لاختبار F لجميع معاملات الميل $\beta_k, \ldots, \beta_2$ تساوي الصفر:

$$H_0 : \beta_2 = \ldots = \beta_k = 0 \tag{3.44}$$

والفرضية البديلة H_1 يكون على الأقل معامل واحد على الأقل من $\beta_2,\ldots,\beta_k$ يختلف عن الصفر $(H_1:$ على الأقل $\beta \neq 0)$، وتحدد احصائية F كما يلي:

$$F(k-1,n-k)=\frac{SSE/(k-1)}{SSR/(n-k)} \tag{3.45}$$

ويتم إجراء اختبار F بمقارنة القيمة بالمستوى الحرج لـ F في العمود $k-1$ درجة حرية، والسطر $n-k$ درجة حرية في جدول F.

يمكن التعبير عن إحصائية F حسب مفهوم R^2 بقسمة كل من البسط والمقام في (3.45) على مجموع المربعات الكلي SST، مع ملاحظة أن $\frac{SSE}{SST}$ هي R^2 و $\frac{SSR}{SST}$ هي $(1-R^2)$:

$$\begin{aligned} F(k-1,n-k) &= \frac{SSE/(k-1)}{SSR/(n-k)} \\ &= \frac{\frac{SSE}{SST}\Big/(k-1)}{\frac{SSR}{SST}\Big/(n-k)} \\ &= \frac{R^2/(k-1)}{(1-R^2)/(n-k)} \end{aligned} \tag{3.46}$$

مثال

نستخدم نموذج عجز الحساب الجاري في ميزان المدفوعات الأردني، ونفترض أن عجز الحساب الجاري CA يعتمد على سعر الصرف الحقيقي الفعّال $REER$ والناتج المحلي الإجمالي الأردني GDP ممثلاً للقدرة الإنتاجية:

$$CA = \beta_1 + \beta_2\, REER + \beta_3\, GDP + u \tag{3.47}$$

الفرضية الأساسية لاختبار F لأفضل تقدير هي أن معاملات الميل تساوي صفر.

$$H_0 : \beta_2 = \beta_3 = 0 \tag{3.48}$$

الفرضية البديلة أن أحد المعاملات على الأقل لا يساوي الصفر.

وكانت نتائج الانحدار كما يلي:

Dependent Variable: CA
Method: Least Squares
Date: 11/27/15 Time: 17:09
Sample (adjusted): 1976 2008
Included observations: 33 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	40.01850	23.22102	1.723374	0.0951
REER	-0.877975	0.199372	-4.403708	0.0001
GDP	-0.000105	0.000893	-0.117546	0.9072

R-squared	0.451376	Mean dependent var	-52.19697
Adjusted R-squared	0.414802	S.D. dependent var	21.27598
S.E. of regression	16.27575	Akaike info criterion	8.503738
Sum squared resid	7947.003	Schwarz criterion	8.639784
Log likelihood	-137.3117	Hannan-Quinn criter.	8.549513
F-statistic	**12.34115**	Durbin-Watson stat	1.321309
Prob(F-statistic)	**0.000123**		

في هذا المثال $k-1$ عدد المتغيّرات التفسيرية تساوي 2 $(k-1=3-1)$، و $n-k$ عدد درجات الحرية وتساوي 30، وبسط إحصائية F هو مجموع المربعات المُفَسَّرة مقسوماً على $k-1$، وهي 2 في السطر، والمقام مجموع مربع البواقي مقسوماً على درجات الحرية 30، وبالتالي فإن إحصائية F حسب (3.46) تكون:

$$F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)} = \frac{0.451376/2}{(1-0.451376)/30} = \frac{0.225688}{0.018287} = 12.34$$

كما في الناتج اعلاه، وفي جميع تطبيقات الانحدار تعتبر إحصائية F جزءًا من تشخيص نتائج الانحدار.

القيمة الحرجة لـ $F(2,30)$ تساوي 3.32 وبالتالي نرفض H_0 عند مستوى معنوية 5٪، وهي متوقعة لأن اختبار t لمعلمة المتغيّر $REER$ معنوي وقيمة t مرتفعة، وبالتالي فإنها لا تساوي الصفر.

غالباً ما تكون إحصائية F معنوية عندما تكون إحصائية t لأي متغيّر معنوية، ومن حيث المبدأ قد لا يكون. افرض أنه لدينا نموذج انحدار متعدد موصّف تماماً و R^2 مرتفعة، ستكون إحصائية F مرتفعة المعنوية. وعلى كل حال، إذا كانت المتغيّرات التفسيرية مرتفعة الارتباط ويخضع النموذج للارتباط المتعدد Multicollinearity سيكون الخطأ المعياري لمعاملات الميل مرتفعاً واحصائية t لجميع المعلمات غير معنوية. وفي هذه الحالة، تكون معلمات النموذج لها قوة تفسيرية عالية؛ لكنك لا تستطيع تحديد مساهمة المتغيّرات التفسيرية منفردة.

كذلك عندما نستخدم الطريقة الأخرى لحساب الإحصائية نحصل على نفس النتيجة:

$$F(k-1, n-k) = \frac{SSE/(k-1)}{SSR/(n-k)} = \frac{6538.346/2}{7947.003/30} = 12.34$$

3-5-5- تحليل إضافي للتباين

إلى جانب اختبار المعادلة الكاملة، نستطيع استخدام اختبار F لمعرفة فيما إذا كانت الاسهامات المشتركة لمجموعة متغيّرات جديدة معنوية أم لا؟ افرض أن النموذج الأصلي كما يلي:

$$Y = \beta_1 + \beta_2 X_2 + \cdots + \beta_k X_k + u \qquad (3.47)$$

نريد إضافة $m-k$ متغيّر جديد، إلى مجموع المربعات المُفسّرة SSE_k، ويصبح النموذج كما يلي:

$$Y = \beta_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \beta_{k+1} X_{k+1} + \cdots + \beta_m X_m + u \qquad (3.48)$$

وبمجموع المربعات المُفسّرة SSE_m يكون لديك مجموع المربعات المُفسّرة الإضافية تساوي $SSE_m - SSE_n$ نستخدم $m-k$ درجة حرية إضافية، ونريد أن نرى فيما إذا كانت الزيادة أكبر من الظاهر.

مرة ثانية سنستخدم اختبار F، وحيث أن SSR_m مجموع المربعات غير المُفسّرة في النموذج الثاني يساوي $SST - SSE_m$، و SSR_k مجموع مربع البواقي في النموذج الأول يساوي $SST - SSE_k$، ويكون التحسن في التقدير عندما نضيف متغيّرات جديدة $SSE_m - SSE_k$ تساوي $SSR_k - SSR_m$، وبالتالي تكون إحصائية F المناسبة كما يلي:

$$F(m-k, n-m) = \frac{(SSR_k - SSR_m)/(m-k)}{SSR_m/(n-m)} \qquad (3.49)$$

وتكون الفرضية الأساسية أن المتغيّر الإضافي لا يساهم في المعادلة:

$$H_0 : \beta_{k+1} = \beta_{k+2} = \cdots = \beta_m = 0 \qquad (3.50)$$

جدول (3-3) تحليل تباين المتغيّرات الأصلية ومجموعة المتغيّرات الإضافية

	مجموع المربعات SS	درجات الحرية df	مجموع المربعات مقسومة على درجات الحرية SS/df	إحصائية F
المُفسّر بالمتغيّرات الأصلية	SSE_k	$k-1$	$SSE_k/(k-1)$	$\frac{SSE_k/(k-1)}{SSR_k/(n-k)}$
البواقي	$SSR_k = SST - SSE_k$	$n-k$	$SSR_k/(n-k)$	
المُفسّر بالمتغيّرات الأصلية	$SSE_m - SSE_k = SSR_k - SSR_m$	$m-k$	$(SSR_k - SSR_m)/(m-k)$	$\frac{(SSR_k - SSR_m)/(m-k}{SSR_m/(n-m)}$
البواقي	$SSR_m = SST - SSE_m$	$n-m$	$SSR_m/(n-m)$	

تتوزع إحصائية F بدرجات حرية $m-k$ و $n-m$، ويبيّن الجزء العلوي من الجدول (3-3) تحليل التباين للقوة التفسيرية للمتغيّرات $k-1$ الأصلية، والجزء السفلي يبين المساهمة الحدية المشتركة للمتغيّرات الجديدة.

مثال

سنستخدم مثال عجز الحساب الجاري في الأردن، وتُظهر النتائج أدناه من انحدار CA على $REER$ لحساب مجموع مربعات البواقي SSR التي تساوي 7950.664.

هل يساهم المتغيّر الجديد بالاشتراك مع الأول بزيادة معنوية القوة التفسيرية للنموذج؟ ستيم النظر إلى اختبار t حيث GDP غير معنوي، وتشكيل اختبار F، وتبين أن SSR تساوي 7947.0.

Dependent Variable: CA
Method: Least Squares
Sample (adjusted): 1976 2008
Included observations: 33 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	38.36136	18.15591	2.112886	0.0428
REER	-0.866637	0.171690	-5.047677	0.0000

R-squared	0.451124	Mean dependent var	-52.19697
Adjusted R-squared	0.433418	S.D. dependent var	21.27598
S.E. of regression	16.01477	Akaike info criterion	8.443592
Sum squared resid	**7950.664**	Schwarz criterion	8.534290
Log likelihood	-137.3193	Hannan-Quinn criter.	8.474109
F-statistic	25.47904	Durbin-Watson stat	1.319565
Prob(F-statistic)	0.000019		

الآن سنضيف متغيّراً جديداً هو *GDP*، وتم تقدير المعادلة وظهرت النتائج للمعادلة الجديدة أدناه:

Dependent Variable: CA
Method: Least Squares
Sample (adjusted): 1976 2008
Included observations: 33 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	40.01850	23.22102	1.723374	0.0951
REER	-0.877975	0.199372	-4.403708	0.0001
GDP	-0.000105	0.000893	-0.117546	0.9072

R-squared	0.451376	Mean dependent var	-52.19697
Adjusted R-squared	0.414802	S.D. dependent var	21.27598
S.E. of regression	16.27575	Akaike info criterion	8.503738
Sum squared resid	**7947.003**	Schwarz criterion	8.639784
Log likelihood	-137.3117	Hannan-Quinn criter.	8.549513
F-statistic	12.34115	Durbin-Watson stat	1.321309
Prob(F-statistic)	0.000123		

لتحسين التقدير بإضافة المتغيّر *GDP* الذي يقلل مجموع مربعات البواقي 7947-7950.663 تكون التكلفة هي انخفاض درجة حرية واحدة؛ لأننا أضفنا متغيّراً جديداً، ويكون مجموع مربعات البواقي غير المُفَسَّر (المتبقي) بعد إضافة *GDP* مساوياً 7947، وعدد درجات الحرية المتبقية بعد إضافة المتغيّر الجديد تساوي 33-3=30.

$$F(m-k,n-m)=\frac{(SSR_k-SSR_m)/(m-k)}{SSR_m/(n-m)}$$

$$F(1,30)=\frac{(7950.664-7947)/(3-2)}{7947/30}=0.0138$$

لذا فإن إحصائية F هي 0.01، وقيمة $F(1,30)$ الحرجة عند مستوى معنوية 5٪ تساوي 4.17، وحيث أن القيمة الحرجة أكبر من قيمة F المحسوبة سيتم قبول H_0 ونستنتج أن معامل *GDP* غير المعنوي لا يساهم بزيادة القوة التفسيرية.

3-6- كتابة تقرير نتائج الانحدار

إذا أردنا كتابة تقرير عن نتائج معادلة الانحدار المتعدد نلخصه بما يلي: (أ) يستخدم التقدير المتغيّرات التالية ... (تذكر المتغيّرات بالأسم)، و (ب) يظهر الخطأ المعياري للمعلمات الظاهرة أسفل المعلمات المقدّرة (أو قيمة t لاختبار الفرضية الأساسية) أن المتغيّر المستقل له تأثير على المتغيّر التابع/ أم لا، و (ج) قيمة R^2.

بالنسبة لمعادلة الطلب على النقود التالية:

$$\hat{m}_i^d = \underset{se\ (0.2615)}{8.665} + \underset{(0.02578)}{0.362}\, gdp_i - \underset{(0.0466)}{1.092}\, r_i, \quad R^2 = 0.967$$

نستطيع قراءة أثر تغيّر المتغيّرات التفسيرية على المتغيّر التابع، ونستطيع التنبؤ بقيم المتغيّر التابع بناءً على قيم المتغيّرات التفسيرية المعطاة، ونحتاج لتكوين فترة التقدير إلى حساب الخطأ المعياري لتقدير معلمات المربعات الصغرى، علماً بأن القيم الحرجة لتوزيع t هي حوالي 2 تقريباً (أو بالتحديد 1.96)، وبالتالي نستطيع الحصول على فترة التقدير باحتمال دقة 95٪ بحساب نقاط الأخطاء المعيارية لتقدير المربعات الصغرى من المعادلة أعلاه.

بالمثل، تستخدم قيمة t لاختبار الفرضية الأساسية $H_0: \beta_K = 0$، ونحصل عليها بقسمة معلمة المربعات الصغرى المقدّرة على قيمة خطأها المعياري، فإذا كانت t أكبر من 2 (تقريباً) ترفض الفرضية الاساسية عند مستوى معنوية $\alpha = 0.05$.

كما أن قيمة R^2 بلغت 0.967 مشيرة إلى أن الناتج المحلي الإجمالي وسعر الفائدة يفسران 96.7٪ من التغيّرات في المتغيّر التابع وهو الطلب على النقود، وأن النسبة المتبقية 3.3٪ تعود لعوامل أخرى.

3-7- تحديد شكل النموذج

قد يتبادر إلى أذهاننا تساؤل عن أفضل طريقة لتقدير معلمات نموذج الانحدار؟ وكيفية تكوين فترات الثقة لمعلمات النموذج المقدّرة؟ وما هي خصائص المعلمات؟ وجميع هذه الأسئلة تتطلب معرفة طبيعة النموذج،

ومن الطبيعي أن نتساءل عن كيفية تحديد النموذج. وهنا سيتم التركيز على الأسئلة التالية: ما هي اعتبارات اختيار النموذج؟ وما هي نتائج اختيار نموذج خاطئ؟ وهل هناك طرق كافية لتقييم النموذج؟

هناك ثلاث صفات جوهرية لاختيار النموذج هي: (1) اختيار شكل الدالة، (2) اختيار المتغيّرات التفسيرية في النموذج، (3) تحقيق نموذج الانحدار المتعدد للفرضيات الثمانية، وسيتم بحث الارتباط الخطي المتعدد، واختلاف التباين، والارتباط الذاتي عند بحث المشاكل القياسية في الفصول اللاحقة، وذلك لاختيار شكل الدالة والمتغيّرات المستقلة والمبادئ الاقتصادية والأسباب المنطقية التي تلعب بصورة بارزة دوراً أساسياً في التحليل. كما نتساءل عن المتغيّرات المؤثرة في المتغيّر التابع Y؟ وكيفية استجابة Y لتغيّر تلك المتغيّرات؟ هل هي بمعدل ثابتٍ؟ أم بمعدلٍ متناقصٍ؟ وهل من المعقول أن نفترض مرونات ثابتة للنموذج الكلي؟ وإجابة تلك الأسئلة يكوّن نقطة الارتكاز لاختيار المتغيّرات المستقلة وشكل الدالة المناسب.

3-7-1 - المتغيّرات المحذوفة

افرض أنك نسيت تضمين المعادلة بأحد المتغيّرات المستقلة المتصلة بها عند وصفها لأول مرة، أو افرض أنك لم تحصل على بيانات عن أحد المتغيّرات؛ ستكون النتيجة في كلا الحالتين **متغيّر محذوف** Omitted variable الذي يُعرّف بأنه متغيّر تفسيري مُهمّ تم استبعاده من معادلة الانحدار، وبالتالي يُصبح لديك متغيّر مستبعد ويصبح تفسير واستخدام المعادلة المقدّرة فيه نظر؛ لأن استبعاد متغيّر مستقل مثل السعر من معادلة الطلب سوف يمنعك من الحصول على تقدير معامل السعر ويسبب تحيّزاً في المعاملات المقدّرة للمتغيّرات الداخلة في المعادلة.

ويسمى التحيّز بسبب استبعاد متغيّر من المعادلة بتحيّز المتغيّرات المحذوفة أو تحيّز التوصيف في معادلة تتضمن أكثر من متغيّر مستقل، وتوضح المعاملات β_k التغير في المتغيّر التابع Y بسبب زيادة وحدة واحدة في المتغيّر المستقل X_k مع بقاء المتغيّرات المستقلة الأخرى في المعادلة ثابتة، فإذا حُذف أحد المتغيّرات عندها لا يكون متغيّراً مستقلاً ولا يكون متغيّراً محدداً في حساب وتفسير $\hat{\beta}_k$، وهذا الحذف قد يسبب تحيّزاً قد يدفع القيمة المتوقعة للمعاملات المقدّرة بالابتعاد عن القيمة الصحيحة لمعلمات المجتمع.

قد نواجه مشكلة اتخاذ قرار بإضافة أو حذف متغيّر تفسيري أو أكثر من النموذج المقدّر، لهذا يُستخدم اختبار t كمعيار آمن للفحص عند تضمين النموذج بأحد المتغيّرات، لكننا سنحتاج عند تضمين النموذج بمجموعة من المتغيّرات الإضافية إلى تقييم تأثير هذا المزيج آخذين بعين الاعتبار هذا النموذج:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \cdots + \beta_k X_{kt} + u_t \quad (3.51)$$

$$Y_t = \beta_1 + \beta_2 X_{2t} + \cdots + \beta_k X_{kt} + \beta_{k+1} X_{k+1t} + \cdots + \beta_m X_{mt} + \varepsilon_t \quad (3.52)$$

ولدينا في هذه الحالة نموذج مقيّد، ونموذج غير مقيّد بمتغيّرات عددها $n-k$ متغيّر لتقييم مزيج الأثر. وتقول الفرضية الأساسية $\beta_{k+1} = \beta_{k+2} = \cdots = \beta_m = 0$ بأن المعنوية المشتركة للمتغيّرات المحذوفة تساوي الصفر، وتقول الفرضية البديلة بأن النموذج (3.52) هو نموذج أساسي ونريد اختبار المتغيّرات المضافة $X_{k+1} = X_{k+2} = \cdots = X_m$ إلى هذا النموذج، ونستخدم اختبار F أو اختبار **Likelihooh Ratio (LR) test**، ويعتمد اختبار F على قيم SSR المختلفة للانحدار المقيّد وغير المقيّد.

مثال

حان الوقت لاكتساب خبرة اختيار المتغيّرات المستقلة، وعليك اتخاذ قرار وصف معادلتك، وللبدء عليك العمل على وصف المعادلة، ولجعله بسيطاً قدّر الإمكان افرض أنك تريد دراسة أثر الصادرات على النمو الاقتصادي والحصول على بيانات عن المتغيّرات المبينة أدناه، والسؤال عن اختيار الوصف.

GDP	الناتج المحلي الإجمالي
g	النمو الاقتصادي
M	المستوردات
FDI	الاستثمار الأجنبي المباشر
Manu	الإنتاج الصناعي
L	العمل
K	رأس المال

افرض أن g هو المتغيّر التابع، وأي متغيّرات مستقل سيتم اختيارها في النموذج؟ وقبل الإجابة فكّر حول المتغيّرات الممكن إضافتها، وماذا تخبرنا الادبيات الاقتصادية حول هذا الموضوع؟ ما هي الإشارات المتوقعة لكل معامل؟ ما هو الأساس النظري الذي يقف خلف كل متغيّر؟ ما هي المتغيّرات الضرورية؟ ما هي المتغيّرات الزائدة؟ هل هناك متغيّرات أخرى سيتم مناقشتها؟

لإخراج هذا المثال إلى حيز التنفيذ عليك أخذ الوقت الكافي لكتابة الوصف الدقيق الذي تريد تنفيذه.

$$g = f(?,?,?,?,?) + \varepsilon$$

من الصعب على أغلب القياسيين المبتدئين تجنب محاولة تضمين جميع المتغيّرات أعلاه في معادلة g واستبعاد أي متغيّر تكون قيمة t له غير

معنوية، وأغلب المبتدئين لا يثقوا بحكمهم ويميلوا لتضمين الكثير من المتغيّرات، وهل تريد إجراء أي تغيير في وصفك المقترح؟ والنتيجة يكون الوصف كما يلي:

$$g = f(\overset{+}{\ln} L, \overset{+}{\ln} K, \overset{+}{\ln} X) + \varepsilon$$

فإذا قدّرنا هذا الوصف باستخدام 28 مشاهدة نحصل على:

$$\hat{g} = 6.91 + \underset{t=(0.01)}{0.0009} \text{ lnK} + \underset{(5.6)}{1.13} \text{ lnL} + \underset{(3.8)}{0.26} \text{ lnX}$$
$$N = 28 \qquad SSR = 0.145813$$

نحن نفضل هذا الوصف لاعتماده على الدراسات التطبيقية وتطابق المعلمات لتوقعاتنا للإشارة، والحجم والمعنوية لدولة نامية، ونعتبر هذه المعادلة مقبولة، ونأخذ ظروف أي تقدير يعتمد على النظرية بعين الاعتبار؛ مثل دالة كوب- دوغلاس، وسيتم حذف متغيّر مهمّ وهو الصادرات X:

$$\hat{g} = 8.36 + \underset{t=(0.92)}{0.078} \text{ lnK} + \underset{(11.5)}{1.75} \text{ lnL}$$
$$N = 28 \qquad SSR = 0.237050$$

لبيان أهمية إضافة متغيّر الصادرات سنجري اختبار F للانحدار المقيّد (المعادلة السابقة) وغير المقيّد (كما في المعادلة الأولى أعلاه) كما يلي:

$$F(m-k, n-m) = \frac{(SSR_k - SSR_m)/(m-k)}{SSR_m/(n-m)}$$

$$F(1, 25) = \frac{(0.237050 - 0.145813)/(3-2)}{0.145813/(28-3)}$$
$$= \frac{0.091237}{0.00583252} = 15.6428$$

وبناءً عليه، وبما أن إحصائية F تساوي 15.6428، وقيمة $F(1, 25)$ الحرجة (الجدولية) عند مستوى معنوية 5٪ تساوي 4.24، فإن القيمة الحرجة أقل من قيمة F المحسوبة سيتم رفض H_0 ونستنتج أن معامل X المعنوي يشارك بزيادة القوة التفسيرية.

3-7-2- اختبار خطأ وصف الانحدار RESET

اختبار **Ramsey Regression Specification Error Test (RESET)** هو أحد أشهر معايير وصف المعادلات، وهو اختبار عام يحدد أعظم احتمال للمتغيّرات المحذوفة أو بعض أخطأ الوصف الأخرى لقياس فيما إذا كان تقدير المعادلة المحددة يمكن تحسينه بإضافة الحدود $\hat{Y}^2$ و $\hat{Y}^3$ و $\hat{Y}^4$ أم لا؟

ما هي الفكرة وراء اختبار RESET؟ تعمل الحدود الإضافية كمتغيّرات بديلة (proxy) عن أي متغيّر ممكن (غير معروف) سواءً كان محذوفاً أو غير ضروري، أو أن شكل الدالة غير صحيح؛ فإذا كانت المتغيّرات البديلة تحسن التقدير الكلي للمعادلة الأصلية حسب اختبار F، سيكون لدينا دليل على وجود بعض أشكال الخطأ في وصف المعادلة، فإذا لم يوجد خطأ وصف ستتوقع بأن معاملات الحدود المضافة ستكون غير معنوية ولا تختلف عن الصفر.

لذلك صمم اختبار Regression Specification Error Test (RESET) ليكشف عن المتغيّرات المحذوفة وشكل الدالة الخاطئ، ويتم على النحو التالي:

1- قدّر المعادلة المراد اختبارها باستخدام المربعات الصغرى العادية:

$$\hat{Y}_i = b_1 + b_2 X_{2i} + b_3 X_{3i} \qquad (3.53)$$

2- قدّر قيم تنبؤ $\hat{Y}_i$ من المعادلة (3.53)، وكوّن الحدود $\hat{Y}^2$ و $\hat{Y}^3$ و $\hat{Y}^4$، وأضفها إلى المعادلة (3.53) كمتغيّرات تفسيرية إضافية، وقدّر المعادلة الجديدة باستخدام المربعات الصغرى العادية OLS:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \gamma_1 \hat{Y}_i^2 + \gamma_2 \hat{Y}_i^3 + \gamma_3 \hat{Y}_i^4 + u_i \qquad (3.54)$$

3- قارن تقدير المعادلة (3.53) و (3.54) باستخدام اختبار F؛ ونختبر الفرضية $H_0 : \gamma_1 = \gamma_2 = \gamma_2 = 0$ مقابل $H_1 : \gamma_1 \neq 0$ و/أو $\gamma_2 \neq 0$ و/أو $\gamma_3 \neq 0$ ويعني رفض H_0 أن المعادلتين مختلفتين، ونستطيع الاستنتاج بأن المعادلة (3.53) سيئة التوصيف وأن النموذج الأصلي غير ملائم ويمكن تحسينه. وعدم رفض H_0 يعني أن الاختبار لا يقبل سوء التوصيف، والفلسفة العامة للاختبار هي أنه إذا استطعنا تحسين النموذج بتضمينه قوة تنبؤية يكون النموذج الأصلي غير ملائم.

مثال

كمثال على اختبار Ramsey RESET سنستخدم مثال الطلب على الدجاج لنرى فيما إذا كان RESET يستطيع كشف خطأ الوصف (حذف متغير أسعار لحوم البقر).

1- الخطوة الأولى تقدير المعادلة الأصلية بدون متغيّر أسعار لحوم البقر (P_B):

$$\hat{Y}_i = 27.5 - \underset{t=(-2.95)}{0.42}\, P_C + \underset{(55.0)}{0.27}\, Yd \qquad (3.55)$$

$$\overline{R}^2 = 0.988 \qquad N = 40 \qquad SSR = 164.31$$

حيث أن $\hat{Y}_i$ الطلب على الدجاج، و P_C أسعار الدجاج، و Yd الدخل المتاح.

2- الخطوة الثانية نحسب $\hat{Y}_i$ من المعادلة (3.55)، ومنها نحسب الحدود $\hat{Y}^2$ و $\hat{Y}^3$ و $\hat{Y}^4$، ثم نعيد تقدير المعادلة (3.55) مع الحدود الثلاثة المضافة:

$$\hat{Y}_i = 243.8 - \underset{t=(-6.39)}{6.3}\, P_C + \underset{(6.35)}{4.2}\, Yd - \underset{(-5.84)}{0.41}\, \hat{Y}_i^2 + \underset{(5.73)}{0.005}\, \hat{Y}_i^3$$

$$+ \underset{(5.61)}{0.00002}\, \hat{Y}_i^4$$

$$\overline{R}^2 = 0.994 \qquad N = 40 \qquad SSR = 79.27 \qquad (3.56)$$

3- في الخطوة الثالثة نقارن تقدير المعادلتين باستخدام اختبار F، وبالتحديد سنختبر فرضية معاملات الحدود الثلاثة المضافة جميعها يساوي الصفر:

$H_0 : \gamma_1 = \gamma_2 = \gamma_2 = 0$

$H_1 :$ غير ذلك

اختبار F المناسب هو ما عرضناه في (3.49).

$$F(m-k, n-m) = \frac{(SSR_k - SSR_m)/(m-k)}{SSR_m/(n-m)} \qquad (3.57)$$

حيث أن SSR_k هو مجموع مربع بواقي المعادلة المقيدة (3.55)، و SSR_m هو مجموع مربع بواقي المعادلة غير المقيدة (3.56)، و $(m-k)$

عدد القيود (3) و $(n-m)$ عدد درجات الحرية في المعادلة غير المقيّدة (64).

$$F(6-3, 40-6) = \frac{(164.31-79.27)/(6-3)}{79.27/(40-6)} = 12.16 \qquad (3.58)$$

قيمة F الحرجة تساوي 2.76 عند مستوى معنوية 5٪ و 3 درجات حرية للبسط و 64 درجة حرية للمقام؛ وبما أن 12.16 أكبر من 2.76 نستطيع رفض الفرضية الأساسية القائلة بأن جميع معاملات المتغيّرات المضافة المشتركة تساوي الصفر، ومنها نستنتج وجود خطأ توصيف في المعادلة (3.55)، وهذه النتيجة لا تدهشنا لأننا نعلم أن أسعار لحوم البقر أهملت من المعادلة، ويخبرنا هذا الاختبار وجود خطأ الوصف لكنه لا يحدد تفاصيل الخطأ.

3-7-3- معيار Akaike و Schwarz

معيار معلومات أكايك Akaike's Information Criterion (AIC) ومعيار شوارتز Schwarz Criterion (SC) هما أسلوبان لمقارنة بدائل التوصيف باستخدام SSR المصحح لحجم العينة (N) وعدد معلمات المعادلة بما فيه الحد الثابت k، ونستطيع استخدام هذه المعايير لتوسيع معايير وصف المعادلة الأساسية عندما نقرر فيما إذا كان التقدير قد تحسن بسبب المتغيّرات الإضافية أم لا، ولها فائدة في تخفيض درجات الحرية وزيادة التعقيد بسبب الإضافة، ومعادلتي الاختبار هما:

$$AIC = \log\left(\frac{SSR}{N}\right) + \frac{2(k)}{N} \qquad (3.59)$$

$$SC = \log\left(\frac{SSR}{N}\right) + \frac{\log(N)(k)}{N} \qquad (3.60)$$

لاستخدام AIC و SC قدّر الشكلين البديلين واحسب AIC و SC لكل معادلة، فإذا انخفضت قيمة AIC و SC يكون الوصف أفضل، لاحظ أن كلا المعيارين يميل إلى معاقبة المتغيّرات الإضافية الأخرى المضافة أكثر من $\bar{R}^2$.

لتطبيق معيار معلومات أكايك ومعيار شوارتز على مثال الطلب على الدجاج، سنرى فيما إذا كان AIC و SC يستطيع اكتشاف خطأ الوصف، ونحن نعلم من المعادلة أنه تم اغفال سعر لحوم البقر، ونحتاج حساب AIC و SC للمعادلة المقيّدة بدون أسعار لحم البقر؛ وتكون المعادلة التي تخفض قيم AIC و SC مع بقاء العوامل الأخرى ثابتة، تكون مواصفاتها مفضلة لدينا.

نموذج الطلب على الدجاج الأصلي الكامل:

$$\hat{Y}_i = 27.6 - \underset{t=(-3.86)}{0.61} P_C + \underset{(2.31)}{0.09} P_B + \underset{(22.07)}{0.24} Yd \qquad (3.61)$$

$$\bar{R}^2 = 0.990 \qquad N = 40 \qquad SSR = 143.07$$

نربط أرقام المعادلة (3.61) بالمعادلة (3.59) و (3.60) ونرى أن:

$$AIC = \log\left(\frac{143.07}{40}\right) + \frac{2(4)}{40} = 1.47$$

$$SC = \log\left(\frac{143.07}{40}\right) + \frac{\log(40)(4)}{40} = 1.64$$

أما المعادلة (3.55) المقيّدة التي حذف منها سعر لحوم البقر كان $SSR = 164.31$ بمتغيّرين مستقلين:

$$AIC = \log\left(\frac{164.31}{40}\right) + \frac{2(3)}{40} = 1.56$$

$$SC = \log\left(\frac{164.31}{40}\right) + \frac{\log(40)(3)}{40} = 1.69$$

بالنسبة لمعيار AIC فقد كان $1.47 < 1.56$ و SC فقد كان $1.64 < 1.69$؛ وبالتالي فكل من المعيارين يعطينا دليلاً على أن المعادلة (3.61) هي أفضل من المعادلة (3.55)، وبالتالي فإن سعر لحوم البقر يخص المعادلة، وهذه الحسابات من الناحية العملية ليست ضرورية لأن أغلب برامج الانحدار تحسب AIC و SC تلقائياً مثل EViews و Stata.

تمارين

3-1- ما هو معنى كل من المصطلحات التالية:

أ- حد الخطأ العشوائي.

ب- التوزيع الطبيعي المعياري.

ج- $se(\hat{\beta})$

د- المقدّر غير المتحيّز.

ه- تقدير BLUE

3-2- قدّر معادلة دالة الاستهلاك، بانحدار الاستهلاك على الناتج المحلي الاجمالي والرقم القياسي للأسعار، واحسب احصائية F باستخدام مجموع المربعات المُفَسَّرة ومجموع البواقي من تقدير الانحدار، وتحقق من تطابق إحصائية F مع نتائج التقدير، ونفذ اختبار القوة التفسيرية للمعادلة كلها، واحسب إحصائية F باستخدام R^2.

3-3- أي من هذه الأزواج للمتغيّرات المستقلة ينتهك الفرضيات الثمانية:

أ- الاستهلاك والدخل المتاح.

ب- X و 2X

ج- X و X^2

د- مقاس الحذاء الأيمن ومقاس الحذاء الأيسر.

3-4- افترض وجود نموذج الانحدار المتعدد التالي:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

ولدينا تسع مشاهدات للمتغيّرات Y_i و X_{1i} و X_{2i} و X_{3i} التالية:

Y_i	X_{1i}	X_{2i}	X_{3i}
1	1	0	1
2	1	1	-2
3	1	2	1
-1	1	-2	0
0	1	1	-1
-1	1	-2	-1
2	1	0	1
1	1	-1	1
2	1	1	0

استخدم الحساب اليدوي للاجابة عن الأسئلة التالية:

أ) احسب قيم المشاهدات حسب مفهوم الخطأ عن الوسط الحسابي، أي أن:

$$x_{2i} = X_{2i} - \bar{X}_2 \quad y_i = Y_i - \bar{Y}, \quad x_{3i} = X_{3i} - \bar{X}_3 ,$$

ب) احسب $\sum y_i x_{2i}$ و $\sum x_{2i}^2$ و $\sum y_i x_{3i}$ و $\sum x_{2i} x_{3i}$ و $\sum x_{3i}^2$.

ج) جد معلمات المربعات الصغرى b_1 و b_2 و b_3.

د) جد بواقي المربعات الصغرى $\hat{u}_1, \hat{u}_2, \cdots, \hat{u}_9$.

هـ) جد تباين التقدير $\hat{\sigma}^2$.

و) جد الخطأ المعياري للمعلمة b_2.

ز) جد SSE و SST و SSR و R^2.

3-5- استخدم نتائج التمرين (3-4).

أ) احسب فترة الثقة للمعلمة β_2 باحتمال 95٪.

ب) اختبر الفرضية $H_0 : \beta_2 = 1$ مقابل الفرضية البديلة $H_1 : \beta_2 \neq 1$.

3-6- عند استخدام مشاهدات عددها $N=40$ مشاهدة لتقدير النموذج التالي:

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 Z_i + u_i$$

حصلنا على $SSR = 979.830$ و $\hat{\sigma}_Y = 13.452229$. جد ما يلي:

أ) R^2.

ب) قيمة احصائية F لاختبار الفرضية الأساسية $H_0 : \beta_2 = \beta_3 = 0$، وهل ترفض الفرضية H_0 أم نفشل في رفضها.

3-7- لديك تقدير معادلة الانحدار التالي (الخطأ المعياري بين الأقواس):

$$\hat{Y} = -120 + \underset{(0.05)}{0.10} F + \underset{(1.00)}{5.33} R, \qquad R^2 = 0.5$$

حيث أن:

Y: ناتج الذرة (شوال/ ايكر) في السنة

F: كثافة السماد (باوند/ ايكر) في السنة

R: كمية المطر (إنش) في السنة

أ) أكتب معنى المعامل 0.10 و 5.33 في هذه المعادلة؛ مبيناً تأثير F و R على Y.

ب) هل الحد الثابت 120- يعني مقداراً سالباً للذرة؟ فإذا لم يكن هذا، فما هو معنى هذا التقدير؟

ج) افرض أنه تم اعلامك بأن القيمة الحقيقية للمعلمة $\beta_2 = 0.20$ ، فهل هذا التقدير متحيّز؟ لماذا أو لماذا لا؟

د) افرض أنك علمت بأن المعادلة لا تحقق الفرضيات الكلاسيكية، وبالتالي هي ليست BLUE، فهل هذا يعني أن β_R الصحيحة لا تساوي 5.33؟ لماذا ولماذا لا؟

الفصل الرابع
النماذج غير الخطية

تعتبر العلاقة غير الخطية أكثر شيوعاً من العلاقات الخطية في العمليات الاقتصادية، وستتعرف في هذا الفصل على معنى تحليل الانحدار الخطي، ونعرض بعض الطرق الشائعة لتقدير العلاقات غير الخطية.

4-1- النماذج الخطية وغير الخطية linearity and nonlinearity

عندما نستخدم مصطلح "تحليل الانحدار الخطي" لا نعرف ماذا نعني بالضبط بالخطية، ومن الضروري تعريفه آخذين بالاعتبار النموذج التالي:

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + u \qquad (4.1)$$

هذا نموذج خطي بمفهومين: أنه خطي في متغيّراته **linear in variables**؛ لأن كل حد يتكون من متغيّر مضروب في معامل parameter، وكذلك خطي في معلماته **linear in parameters**؛ لأن كل حد يتكون من معلمات مضروبة في متغيّر.

ولأغراض تحليل الانحدار الخطي، فالمفهوم الثاني هو الأهم، لمقدرتنا تجنب عدم الخطية في المتغيّرات باستخدام التعريف المناسب له، مثلاً على فرض أن العلاقة كانت على الشكل التالي:

$$Y = \beta_1 + \beta_2 X_2^2 + \beta_3 \sqrt{X_3} + \beta_4 \log X_4 + u \qquad (4.2)$$

يمكن تعريف $Z_2 = X_2^2$ و $Z_3 = \sqrt{X_3}$ و $Z_4 = \log X_4$ ونعيد كتابة العلاقة كما يلي:

$$Y = \beta_1 + \beta_2 Z_2 + \beta_3 Z_3 + \beta_4 Z_4 + u \qquad (4.3)$$

أصبحت الآن المتغيّرات خطية كما هي المعلمات، وهذا النوع من التحويل هو تجميلي فقط، كم نعرض معادلة الانحدار بشكلها الأصلي غير الخطي، لكي نتجنب الحاجة لأي توضيح بوضع الملاحظات الاضافية. ومن جهة أخرى، نعرض المعادلة غير خطية في كل من المعلمات والمتغيّرات التالية:

$$Y = \beta_1 X_2^{\beta 2} \qquad (4.4)$$

سنبدأ بمثال نموذج بسيط، وهو شكل الدالة المعكوس؛ الذي يعبّر عنه بأن Y دالة في معكوس أحد المتغيّرات المستقلة أو أكثر (مثل X_2):

$$Y = \beta_1 + \beta_2 \frac{1}{X_2} + \beta_3 X_3 + u \qquad (4.5)$$

ويستخدم شكل الدالة المعكوس عندما يكون الأثر المتوقع لمتغير مستقل معين يقترب من الصفر؛ وذلك عندما تقترب قيمته من ما لا نهاية؛ مثلاً عندما تكون X_2 كبيرة يكون أثره على Y منخفضاً.

ولا يساوي X_2 في المعادلة (4.5) الصفر، وإذا كان مساوياً للصفر يكون حاصل القسمة عليه مساوياً لقيم غير معرّفة ويكون الميل بالنسبة إليه كما يلي:

$$\frac{\Delta Y}{\Delta X_2} = \frac{-\beta_2}{X_2^2} \tag{4.6}$$

ويكون ميل X_2:

1- عندما تكون β_2 موجبة يكون الميل بالنسبة إلى X_2 سالباً، وتتناقص بالقيمة المطلقة عندما يزداد X_2، وتقترب نتيجة العلاقة بين Y و X_2 مع بقاء X_3 ثابتاً من $\beta_1 + \beta_3 X_3$ عندما يزداد X_2 (مع تجاهل حد الخطأ).

2- عندما تكون β_2 سالبة، ستقطع العلاقة محور X_2 عند $-\beta_2 /(\beta_1 + \beta_3 X_3)$، وتميل للأعلى تجاه الخط الأفقي (يسمى خط متقارب) عندما تكون β_2 موجبة.

وتتواجد تطبيقات الشكل المعكوس في عدة حقول في النظرية الاقتصادية وفي العالم الحقيقي، ومنها منحنى فيليبس **Phillips curve**، للعلاقة غير الخطية بين معدل البطالة ونسبة التغيّر في الأجور، نفترض أن نسبة تغيّر الأجور W علاقتها سالبة بمعدل البطالة U، وهذه الفرضية تختبر بشكل الدالة المعكوسة:

$$W = \beta_1 + \beta_2 \frac{1}{U} + u \tag{4.7}$$

تم تقدير المعادلة باستخدام OLS، وحصلنا على تقدير يخص الاقتصاد الأمريكي كما يلي:

$$\hat{W} = 0.00679 + \underset{\substack{(0.0590) \\ t=3.20}}{0.1842} \frac{1}{U} \qquad (4.8)$$

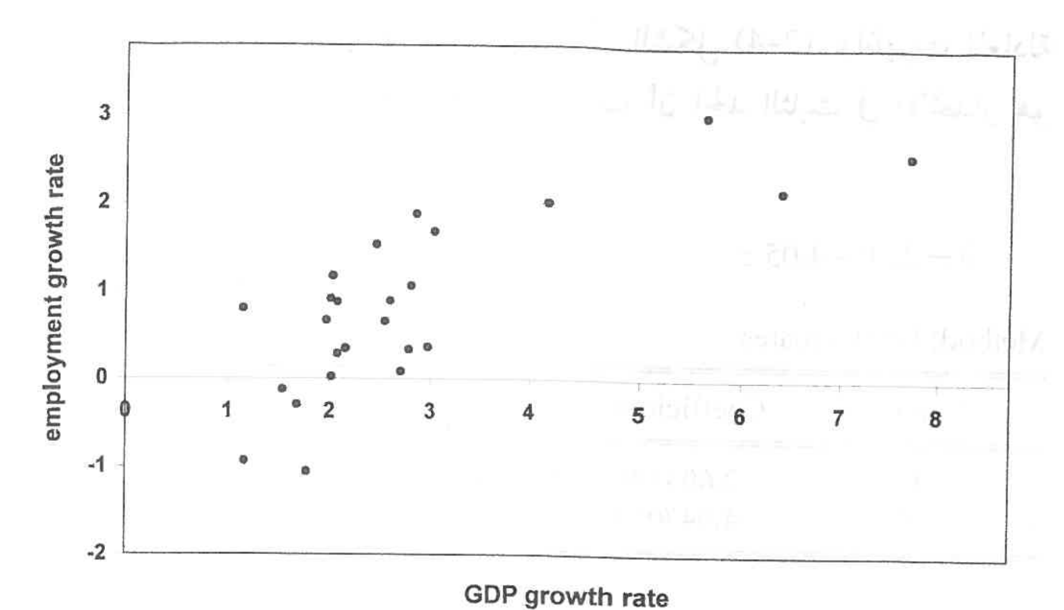
$R^2 = 0.397$

وهذا يشير إلى أن W و U مرتبطتين حسب الحالة (أ) أعلاه.

كما استخدمنا بيانات الجدول (4-1) الذي يتضمن متوسط معدل النمو السنوي للعمالة والناتج المحلي الإجمالي لخمس وعشرين دولة من دول OECD التي يعرضها الشكل (4-1).

جدول (4-1) معدل النمو السنوي المتوسط للعمالة (e) والناتج المحلي الإجمالي (g)، خلال الفترة 1988-1997

متوسط معدلات النمو السنوي							
	e	(g)	Z=1/g		e	(g)	Z=1/g
Australia	1.68	3.04	0.329	Korea	2.57	7.73	0.129
Austria	0.65	2.55	0.392	Luxembourg	3.02	5.64	0.177
Belgium	0.34	2.16	0.463	Netherlands	1.88	2.86	0.350
Canada	1.17	2.03	0.493	New Zealand	0.91	2.01	0.498
Denmark	0.02	2.02	0.495	Norway	0.36	2.98	0.336
Finland	–1.06	1.78	0.562	Portugal	0.33	2.79	0.358
France	0.28	2.08	0.481	Spain	0.89	2.60	0.385
Germany	0.08	2.71	0.369	Sweden	–0.94	1.17	0.855
Greece	0.87	2.08	0.481	Switzerland	0.79	1.15	0.870
Iceland	–0.13	1.54	0.649	Turkey	2.02	4.18	0.239
Ireland	2.16	6.40	0.156	UK	0.66	1.97	0.508
Italy	–0.30	1.68	0.595	United States	1.53	2.46	0.407
Japan	1.06	2.81	0.356				



شكل رقم 4-1: معدلات نمو العمالة ونمو الناتج المحلي الإجمالي

حيث يتضح من الشكل أن هذه العلاقة غير خطية، وسنحدد الشكل التالي للعلاقة غير الخطية بالنموذج التالي:

$$e = \beta_1 + \frac{\beta_2}{g} + u \qquad (4.9)$$

هذه العلاقة غير خطية في g، إلا أننا نستطيع اعادة كتابة النموذج ليصبح خطياً في المتغيّرات كما هو في المعلمات إذا عرّفنا $z = \frac{1}{g}$:

$$e = \beta_1 + \beta_2 z + u \qquad (4.10)$$

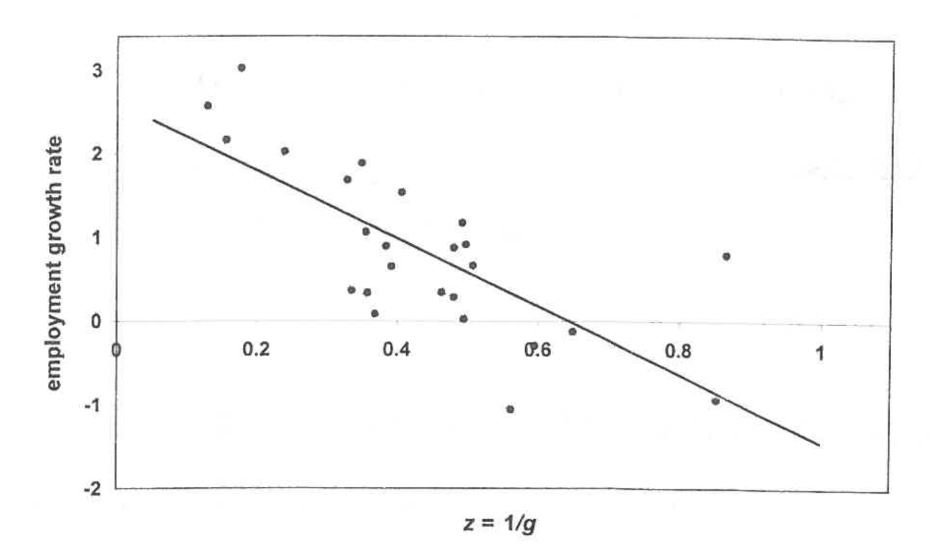
وحسبت بيانات z حسب الصيغة $z = \frac{1}{g}$ كما في الجدول (4-1) من عمود g لاستخدامها في تطبيقات الانحدار، وأظهرت نتائج انحدار e

على z أدناه، وتم رسم الانحدار كما في الشكل (4-2)، وأظهرت المعادلة (4.11) نتائج هذا الانحدار؛ حيث بيّنت أن الحد الثابت في الانحدار هو تقدير β_1، ومعلمة z هي تقدير β_2:

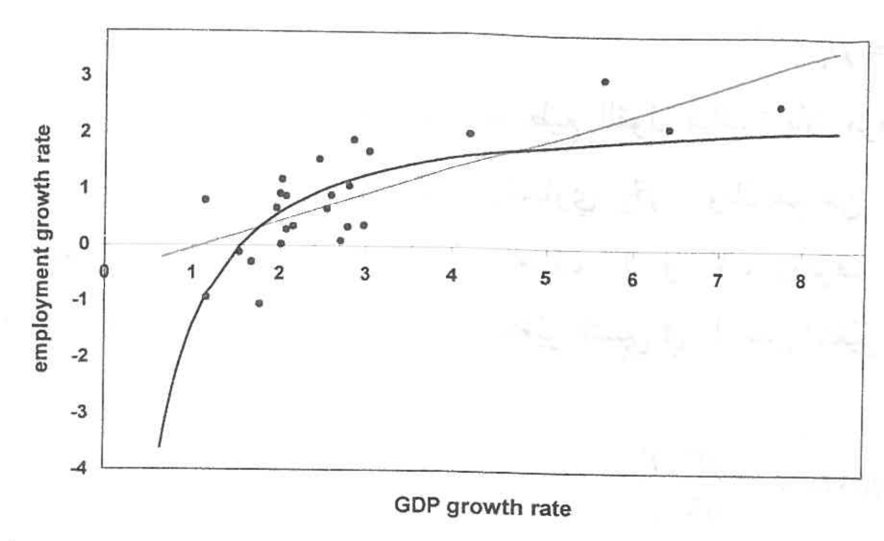
$$\hat{e} = 2.60 - 4.05\, z \qquad (4.11)$$

Method: Least Squares

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.603440	0.374835	6.945569	0.0000
Z	-4.047014	0.793267	-5.101708	0.0000



شكل (4 -2) انحدار معدل نمو العمالة على مقلوب معدل نمو الناتج المحلي الاجمالي



شكل (4-3) الانحدار الخطي وغير الخطي لمعدل نمو العمالة e على معدل نمو الناتج المحلي الاجمالي g

ثم نعوض $z=\frac{1}{g}$ تصبح:

$$\hat{e}=2.60-4.05z=2.60-\frac{4.05}{g} \qquad (4.12)$$

يُظهر الشكل (4-3) العلاقة غير الخطية (4.12) والتي رسمت في شكلها الأصلي. وكان من الواضح في هذه الحالة أن العلاقة بين e و g غير خطية، وفي حالة نحصل على العلاقة غير الخطية باستخدام طريقة الرسم.

4-2- التحويل اللوغاريتمي

4-2-1- النماذج اللوغاريتمية

سنعالج الدالة (4.4) غير الخطية في معلماتها كما هو الحال في متغيّراتها كما يلي:

$$Y = \beta_1 X^{\beta_2} \quad (4.13)$$

عندما ترى مثل هذه الدالة تستطيع القول مباشرة بأن مرونة Y بالنسبة للمتغيّر X هي قيمة ثابتة وتساوي β_2 . وبالرغم من العلاقة الرياضية التي تربط Y و X أو تعرّف Y و X، وتعرّف مرونة **(elasticity)**[1] Y بالنسبة للمتغيّر X بالتغيّر النسبي في Y على التغيّر النسبي في X:

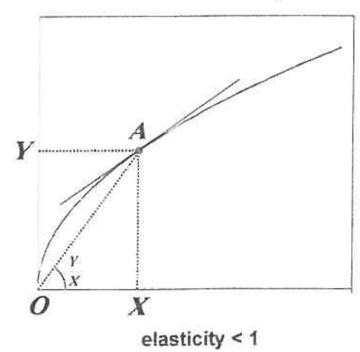
$$elasticity = \frac{dY/Y}{dX/X} \quad (4.14)$$

إذا كان Y الطلب، و X الدخل؛ فإنها تعرّف بمرونة الطلب الدخلية على السلع. وقد نعيد كتابة الصيغة كما يلي:

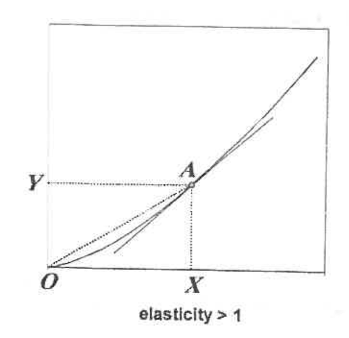
$$elasticity = \frac{dY/dX}{Y/X} \quad (4.15)$$

1

$$\text{المرونة} = \frac{dY/Y}{dX/X} = \frac{dY/dX}{Y/X}$$



elasticity < 1



elasticity > 1

يمكن تفسير الطلب بالميل الحدي لاستهلاك البضاعة مقسوماً على معدل الميل للاستهلاك، وإذا كانت العلاقة بين Y و X تأخذ الشكل (4.13) التالي:

$$\frac{dY}{dX} = \beta_1 \beta_2 X^{\beta_2 - 1} = \beta_2 \frac{X}{Y} \tag{4.16}$$

تكون المرونة كما يلي:

$$elasticity = \frac{dY/dX}{Y/X} = \frac{\beta_2 Y/X}{Y/X} = \beta_2 \tag{4.17}$$

ومن الأمثلة على ذلك منحنى انجل **Engel curve** التالي:

$$Y = 0.01 X^{0.3}$$

هذا يعني أن مرونة الطلب الدخلية تساوي 0.3، وإذا حاولنا شرح المرونة لأي شخص غير معتاد على اللغة الاقتصادية، فإن أسهل طريقة لشرح المرونة له بالقول أن حدوث تغيّر في الدخل X بنسبة 1٪ سيسبب تغيّراً في الطلب على Y بنسبة 0.3٪.

يمكن تحويل الدالة من هذا النوع إلى معادلة خطية باسلوب التحويل اللوغاريتمي **Logarithmic transformation**، ويُبيّن الصندوق (4-1) خصائص اللوغاريتمات الأساسية.

صندوق (4-1) استخدام اللوغاريتم

بعض القواعد الأساسية:

1- إذا كان $Y = XZ$ فإن $\log Y = \log X + \log Z$.

2- إذا كان $Y = X/Z$ فإن $\log Y = \log X - \log Z$.

3- إذا كان $Y = X^n$ فإن $\log Y = n \log X$.

يمكن مزج هذه القواعد لتحويل صيغ معقدة مثل المعادلة (4.13) التالية:

إذا كانت المعادلة $Y = \beta_1 X^{\beta_2}$ فإن:

- استخدم القاعدة 1 $\log Y = \log \beta_1 + \log X^{\beta_2}$
- استخدم القاعدة 3 $= \log \beta_1 + \beta_2 \log X$

لم نحدد فيما إذا كان اللوغاريتم يأخذ الأساس e أو الأساس 10. وعادة نستخدم الأساس e أو ما يسمى باللوغاريتم "الطبيعي"، وهذا معياري في الاقتصاد القياسي، وفي بعض الأحيان نكتب ln بدلاً من log لنؤكد أننا نستخدم اللوغاريتم الطبيعي، لكنة الآن غير ضروري، ولا يستخدم اللوغاريتم للأساس 10.

يمكن استخدام القاعدة التالية للأساس e.

4- إذا كان $Y = e^X$ فإن $\log Y = X$.

ونكتب e^X وكذلك نكتب $\exp(X)$ وتسمى معكوس X، ويمكن القول بأن $\log e^X$ هي لوغاريتم معكوس X، وبما أن log و antilog يحذفان بعضهما، فمن غير المدهش أن $\log e^X$ تحول لتكون X، وباستخدام القاعدة 2 فإن $\log e^X = X \log e = X$، حيث أن $\log e$ للأساس e تساوي 1.

ويمكن تحويل المعادلة (4.13) معادلة خطية كما يلي:

$$\begin{aligned} \log Y &= \log \beta_1 X^{\beta_2} \\ &= \log \beta_1 + \log X^{\beta_2} \\ &= \log \beta_1 + \beta_2 \log X \end{aligned} \tag{4.18}$$

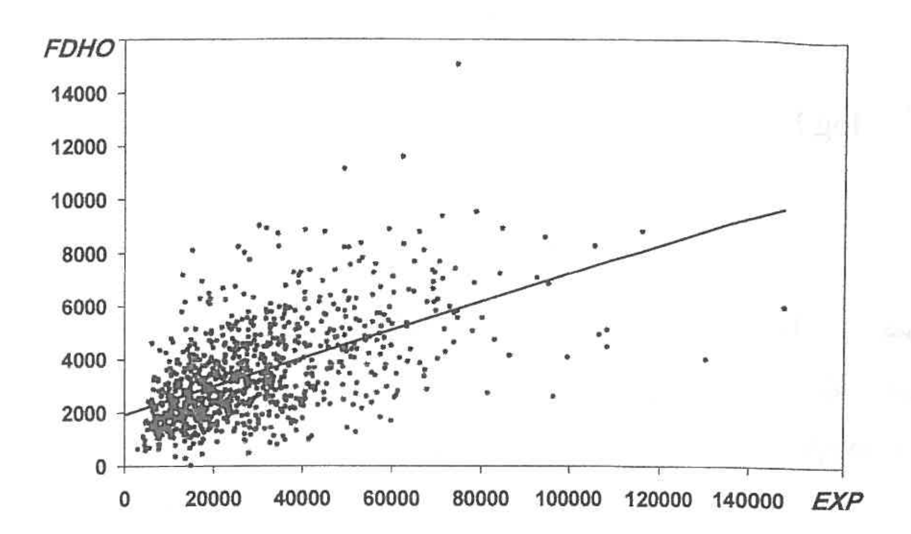
يسمى هذا **بالنموذج اللوغاريتمي logarithmic model** أو **نموذج لوغاريتمي خطي loglinear model**، مشيراً إلى حقيقة أنه خطي باللوغاريتمات، فإذا عبّرنا عن $Y' = \log Y$ و $X' = \log X$ و $\beta_1' = \log \beta_1$ يمكن كتابة المعادلة كما يلي:

$$Y' = \beta_1' + \beta_2 X' \tag{4.19}$$

وتكون إجراءات تقدير الانحدار كما يلي: نحسب أولاً Y' و X' لجميع المشاهدات بأخذ اللوغاريتم للبيانات الأصلية. ثمّ نقدّر انحدار Y' على X' ثانياً، ويقدر لنا معامل X' المعلمة β_2، ويقدر الحد الثابت β_1' وهو $\log \beta_1$، وللحصول على قيمة β_1 الأصلية عليك أخذ معكوس اللوغاريتم antilog والذي يحسب بالصيغة $\exp(\beta_1')$.

مثال: منحنى انجل Engel Curve

يُبيّن الشكل (4-4) انفاق القطاع العائلي على الطعام في المنازل FDHO، ومجموع الانفاق العائلي السنوي بالدولار الأمريكي لـ 869 عائلة في الولايات المتحدة في عام 1995، وأخذت البيانات من مسح نفقات الأسرة Consumer Expenditure Survey.



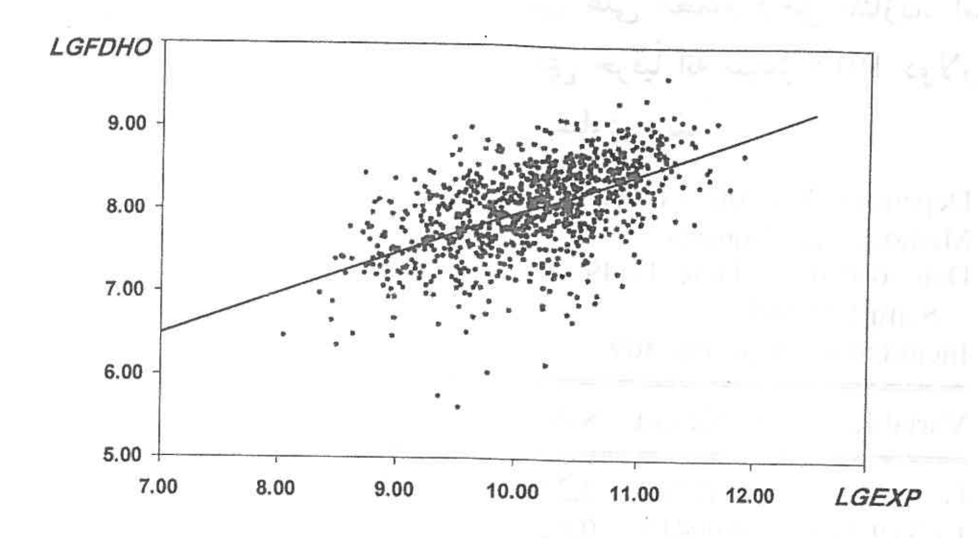
شكل (4-4) انحدار الانفاق على الطعام على اجمالي الانفاق العائلي

عندما نحلل بيانات نفقات العائلات وهي ترتبط بنوع الانفاق إلى اجمالي النفقات العائلية بدلاً من الدخل؛ وتسبب علاقة تجعل النفقات أكثر استقراراً من الدخل، وتبيّن النتائج أدناه ناتج الانحدار الخطي اللوغاريتمي.

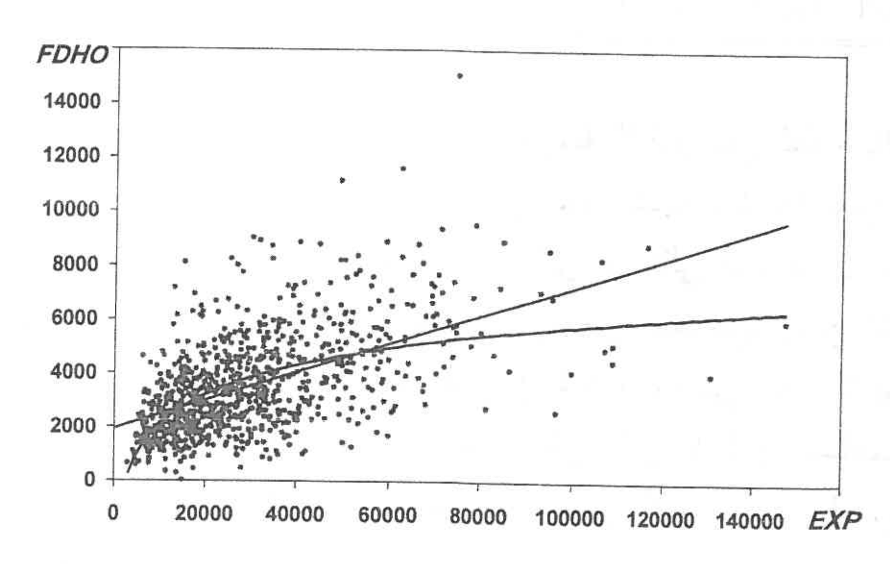
Dependent Variable: FDHO
Method: Least Squares
Sample: 1 869
Included observations: 869

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1916.143	96.54591	19.84696	0.0000
Expend	0.052843	0.002706	19.53124	0.0000

R-squared	0.305550
F-statistic	381.4694
Prob(F-statistic)	0.000000



شكل (4-5) انحدار لوغاريتم الانفاق على الطعام على لوغاريتم اجمالي الانفاق العائلي



شكل (4-6) الانحدار الخطي واللوغاريتمي للانفاق على الطعام على اجمالي الانفاق العائلي

تشير نتائج الانحدار الخطي إلى أن ما نسبته 5.3٪ من قيمة الدولار الحدي (آخر دولار ينفقه الفرد) سينفق على الطعام داخل المنازل. أما تفسير الحد الثابت فهو مشكلة؛ لأنه يعني حرفياً أنه سينفق 1916 دولاراً على الطعام حتى لو كان الانفاق الكلي يساوي صفراً.

Dependent Variable: LFDHO
Method: Least Squares
Date: 09/20/12 Time: 11:49
Sample: 1 869
Included observations: 869

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.166271	0.244297	12.961	0.0000
LEXPEND	0.4800417	0.0241212	19.901	0.0000

R-squared	0.3138
F-statistic	396.06
Prob(F-statistic)	0.000000

كما يُبيّن الانحدار اللوغاريتمي أن مرونة الانفاق على الطعام بالنسبة لاجمالي الانفاق العائلي هي 0.48، فهل هذه النتيجة معقولة؟ نعم؛ لأنه يتم أكل الطعام الضروري بدلاً من الكمالي، وبالتالي سنتوقع أن المرونة تكون أقل من 1، والحد الثابت ليس له أيّ معنى اقتصادي، ويبيّن الشكل (4-6) رسم خط الانحدار اللوغاريتمي في شكله الأصلي، لكنه لا يوجد اختلاف كبير بين خطوط الانحدار فوق الجزء الأوسط لمدى المشاهدات، ومن الواضح أن الانحدار اللوغاريتمي يعطي أفضل تقدير لمستويات الانفاق العائلي المنخفضة جداً والمستويات المرتفعة جداً.

4-2-2- النماذج شبه اللوغاريتمية

من أشكال الدوال الشائعة هو شكل المعادلة التالية:

$$Y = \beta_1 e^{\beta_2 X} \tag{4.20}$$

تفسر β_2 بالتغيّر النسبي في Y لتغيّر وحدة في X، ومن السهل عرض تفاضلها كما يلي:

$$\frac{dY}{dX} = \beta_1 \beta_2 e^{\beta_2 X} = \beta_2 Y \tag{4.21}$$

وبالتالي:

$$\frac{dY/Y}{dX} = \beta_2 \tag{4.22}$$

ومن الناحية التطبيقية، من الطبيعي القول بنسبة التغيّر percentage change في Y بدلاً التغيّر النسبي proportional change لتغيّر وحدة في X، وتقدير β_2 مضروباً في 100.

يمكن تحويل الدالة إلى نموذج خطي في معلماته بأخذ اللوغاريتم للجانبين:

$$\begin{aligned}
\log Y &= \log \beta_1 e^{\beta_2 X} \\
&= \log \beta_1 + \log e^{\beta_2 X} \\
&= \log \beta_1 + \beta_2 X \log e \\
&= \log \beta_1 + \beta_2 X
\end{aligned} \tag{4.23}$$

لاحظ أن اللوغاريتم على الجانب الأيسر للمتغيّرات، ولهذا السبب توصف المعادلة (4.23) بالنموذج شبه اللوغاريتمي **semilogarithimic model**.

تفسر β_2 بالتغيّر النسبي في Y لتغيّر وحدة في X ويكون هذا صالحاً عندما تكون β_2 صغيرة فقط، وإذا كانت β_2 كبيرة قد يكون التفسير أكثر تعقيداً. وعلى فرض أن Y ترتبط في X حسب المعادلة (4.20) وتزيد X بوحدة واحدة بالنسبة لـ X'، فإن Y' القيمة الجديدة من Y تعطى كما يلي:

$$
\begin{aligned}
Y' = Y + \Delta Y &= \beta_1 e^{\beta_2 (X+\Delta X)} \\
&= \beta_1 e^{\beta_2 X} e^{\beta_2 \Delta X} \\
&= Y e^{\beta_2 \Delta X} \\
&= Y\left(1+\beta_2 \Delta X + \frac{(\beta_2 \Delta X)^2}{2} + \ldots\right) \qquad (4.24)
\end{aligned}
$$

بما أن $e^Z = 1 + Z + \frac{Z^2}{2!} + \frac{Z^3}{3!} + \ldots$. فإن:

$$\Delta Y = Y\left(\beta_2 \Delta X + \frac{(\beta_2 \Delta X)^2}{2} + \ldots\right)$$

التغيّر النسبي لتغيّر X بوحدة واحدة هي فعلياً أكبر من β_2. وإذا كانت β_2 صغيرة (أقل من 0.1)؛ فإن β_2^2 وبقية الحدود ستكون صغيرة جداً ونستطيع اهمالها، وفي هذه الحالة يبسط الجانب الأيمن من المعادلة إلى $Y(1+\beta_2)$ وتفسر β_2 الأصلية.

4-2-3- حد الخطأ

كيف يتأثر حد الخطأ بتلك التحويلات؟ يتطلب ظهور حد الخطأ في المعادلة المُحوّلة باضافة الحد $(+u)$ يلبي شروط نموذج الانحدار. فإذا لم يحقق الشروط، فإن معاملات انحدار المربعات الصغرى لا يكون لها الخصائص المعتادة، ويصبح الاختبار غير صالحٍ. مثلاً من المرغوب فيه أن تكون المعادلة (4.9) على الشكل التالي:

$$e = \beta_1 + \beta_2 Z + u \qquad (4.25)$$

عندما نأخذ الأثر العشوائي في الحساب، نأخذ العكس؛ وهذا يعني أن المعادلة الأصلية (غير المُحوّلة) تأخذ الشكل التالي:

$$e = \beta_1 + \frac{\beta_2}{g} + u \qquad (4.26)$$

في هذه الحالة الخاصة، إذا كان حد الخطأ مضافاً إلى المعادلة الأصلية بشكل صحيح يفي بشروط نموذج الانحدار، سيكون صحيحاً في المعادلة المُحوّلة ولا يكون مشكلة هنا.

ماذا سيحدث إذا بدأنا بالنموذج التالي:

$$Y = \beta_1 X_2^{\beta_2} \qquad (4.27)$$

كما نرى أن نموذج الانحدار بعد تحويله إلى نموذج خطي بأخذ اللوغاريتم يصبح كما يلي:

$$\log Y = \log \beta_1 + \beta_2 \log X + u \qquad (4.28)$$

عندما يكون حد الخطأ موجوداً، عد إلى المعادلة الأصلية فهذا يعني أن المعادلة (4.27) يجب أن تُكتب كما يلي:

$$Y = \beta_1 X^{\beta 2} \upsilon \qquad (4.29)$$

حيث أن υ و u مرتبطة حسب $\log \upsilon = u$، وبالتالي الحصول على حد الخطأ بدلاً من الحجم العشوائي. لاحظ أن $u = 0$ عندما $\log \upsilon = 0$ التي تحدث عندما تكون $\upsilon = 1$. وسيساوي العامل العشوائي الصفر في تقدير المعادلة (4.28) حتى وإن ساوى υ للواحد. فإذا ساوت υ للواحد لا يتم تعديل $\beta_1 X_2^{\beta 2}$.

لكي يكون اختبار t و F صحيحاً يجب أن يكون توزيع u توزيعاً طبيعياً؛ وهذا يعني أن $\log \upsilon$ يجب أن يكون توزيعه طبيعياً، وهذا يحدث فقط عندما يكون توزيع اللوغاريتم طبيعي لحد الخطأ υ. ماذا يحدث إذا افترضنا أن حد الخطأ في المعادلة الأصلية كان مضافاً بدلاً من أن يكون مضروباً؟

$$Y = \beta_1 X_2^{\beta 2} + u \qquad (4.30)$$

يكون الجواب أنه عندما نأخذ اللوغاريتم للمعادلة، فإنه لا يوجد طريقة رياضية نستطيع استخدامها لتبسيط المعادلة $\log \left(Y = \beta_1 X_2^{\beta 2} + u\right)$، ولا يؤدي التحويل إلى معادلة خطية، وعليك استخدام أسلوب الانحدار غير الخطي كما سيتم شرحه في المقطع التالي.

تمارين

4-1- حمّل بيانات الملف CES من الأنترنت وقدر انحدار خطي وانحدار لوغاريتمي للسلع على EXP واجمالي الانفاق العائلي واستثني المشاهدات التي انفاقها صفر على السلع، وفسر نتائج الانحدار واستخدم الاختبارات المناسبة.

4-2- أعد تقدير الانحدار اللوغاريتمي في التمرين (4-1) وأضف إليه لوغاريتم حجم العائلة كمتغيّر تفسيري اضافي، وفسر نتائج الانحدار واستخدم الاختبارات المناسبة.

4-3- نماذج تحوي متغيّرات تربيعية وتفاعلية

نأتي الآن على نموذج بحدود تربيعية quadratic مثل:

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_2^2 + u \qquad (4.31)$$

ونموذج بحدود تفاعلية interactive مثل:

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_2 X_3 + u \qquad (4.32)$$

قد يُعرض النموذج التربيعي كحالة خاصة من نموذج تفاعلي مثل $X_3 = X_2$، لكن من المناسب معالجتها كل على انفراد. ويمكن تقدير هذه النماذج باستخدام طريقة المربعات الصغرى بدون تعديل، إلا أن تفسير تلك المعاملات يتم بعرض أثر تغيّر المتغيّر مع ثبات بقية المتغيّرات، إلا أنه

لا يمكن تفسير تغيّر X_2 في حالة النموذج التربيعي بدون تغيّر X_2^2 كذلك، وليس من الممكن تفسير تغيّر X_2 في النموذج التفاعلي بدون تغيّر X_2X_3 كذلك، إذا بقي المتغيّر X_2 ثابتاً.

4-3-1- المتغيّرات التربيعية

باشتقاق (4.31) نحصل على التغيّر في Y لكل وحد تغيّر في X_2:

$$\frac{dY}{dX_2} = \beta_2 + 2\beta_3 X_2 \qquad (4.33)$$

تستطيع رؤية أثر تغيّر وحدة واحدة في X_2 على Y، ويكون حجم التغيّر $(\beta_2 + 2\beta_3 X_2)$ عندما يتغيّر X_2، وهذا يعني أن β_2 تفسر التغيّر في النموذج الخطي العادي:

$$Y = \beta_1 + \beta_2 X_2 + u \qquad (4.34)$$

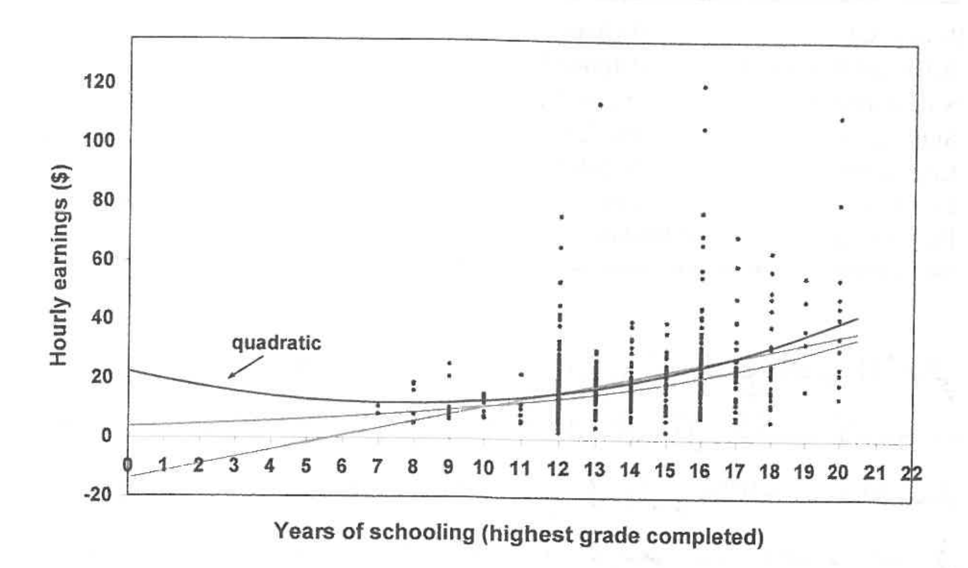
حيث أنه غير تام الأثر لتغيّر وحدة في X_2 على Y، ويفسر β_2 في المعادلة (4.33) أثر تغيّر وحدة في X_2 على Y كحالة خاصة عندما $X_2 = 0$، أما بالنسبة لقيم X_2 غير الصفرية ستكون المعلمة مختلفة.

كذلك تفسر β_3 كتفسير خاص إذا أعيد كتابة النموذج كما يلي:

$$Y = \beta_1 + (\beta_2 + \beta_3 X_2) X_2 + u \qquad (4.35)$$

يمكن تفسير β_3 كمعدل تغيّر معامل X_2 لكل وحدة تغيّر في X_2. ويتم تفسير β_1 تفسيراً تقليدياً، وكالمعتاد تكون قيمة Y جزءاً من المكوّن العشوائي عندما يكون $X_2 = 0$.

يوجد لدينا مشاكل اضافية منها تقدير الحد الثابت؛ قد لا يكون معناه معقولاً إذا كان $X_2 = 0$ ويكون خارج نطاق البيانات. مثلاً في حالة الانحدار الخطي للأجور على التعليم (مقاساً بعدد سنوات التعليم) earnings on schooling reproduced كما في الشكل (4-10) كان الحد الثابت سالباً؛ وهذا يعني أن الأفراد يكسبوا أجرهم بدون تعليم ما مقداره 13.93- دولار لكل ساعة، فإذا كان $X_2 = 0$ تقع خارج نطاق البيانات، ونفس نوع التشويه يمكن أن يحدث عند تقدير β_2 .



شكل (4-10) انحدار الايراد على التعليم التربيعي والخطي وشبه اللوغاريتمي

يُبيّن الجدول أدناه ناتج انحدار الأجور (SSQ مربع الأجور) على التعليم؛ وتعني معلمة S أن أثر سنة تعليم يخفض مكاسب الساعة بـ 1.76 دولار، وكذلك تفسير الحد الثابت غير واقعي، وبالتالي يكون أجر الأفراد بدون تعليم 16.107 دولار لكل ساعة وهو غير قابل للتصديق.

Dependent Variable: EARNINGS
Method: Least Squares
Date: 09/23/12 Time: 12:47
Sample (adjusted): 3 5299
Included observations: 4389 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	16.10735	3.475271	4.634846	0.0000
S	-1.764269	0.503542	-3.503716	0.0005
S^2	0.112028	0.017941	6.244260	0.0000

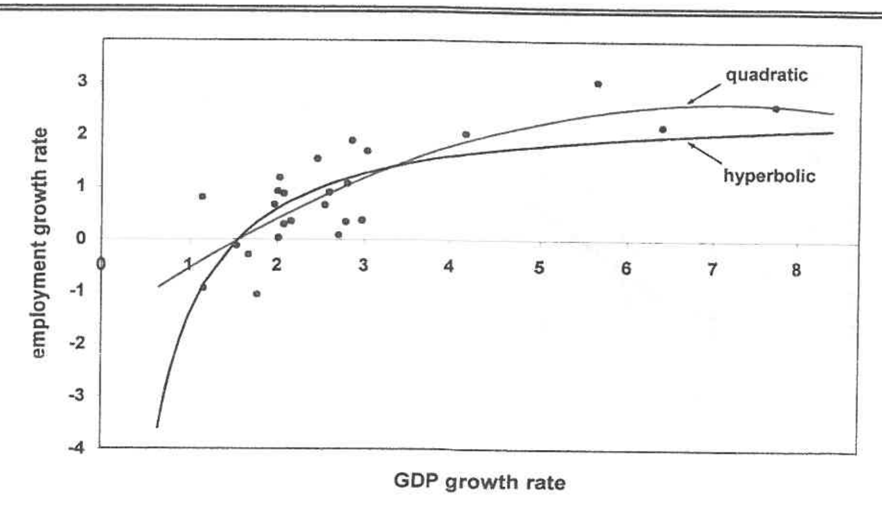
R-squared	0.101054	Mean dependent var	13.25825
Adjusted R-squared	0.100644	S.D. dependent var	10.82848
S.E. of regression	10.26912	Akaike info criterion	7.496844
Sum squared resid	462525.1	Schwarz criterion	7.501209
Log likelihood	-16448.82	Hannan-Quinn criter.	7.498384
F-statistic	246.5231	Durbin-Watson stat	1.889701
Prob(F-statistic)	0.000000		

تزودنا بيانات معدل نمو العمالة e ومعدل نمو الناتج المحلي الاجمالي g لخمس وعشرين دولة من دول OECD بمثال أقل اشكالية لاستخدام الدالة التربيعية، حيث تم تعريف gsq مربع g، وتبيّن النتائج أدناه الانحدار التربيعي، ويقارن الشكل (4-11) الانحدار التربيعي كما حصلت عليه في المقطع (4-1)، ويظهر تحديد المعادلة التربيعي تحسينا لتقدير دالة قطع زائد hyperbolic في المقطع (4-1).

Dependent Variable: E
Method: Least Squares
Sample: 1 25
Included observations: 25

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.678113	0.655664	-2.559410	0.0179
G	1.200205	0.386223	3.107548	0.0051
G^2	-0.083841	0.044569	-1.881133	0.0733

R-squared	0.646850	Mean dependent var	0.833600
Adjusted R-squared	0.614745	S.D. dependent var	1.014519
S.E. of regression	0.629701	Akaike info criterion	2.025023
Sum squared resid	8.723511	Schwarz criterion	2.171288
Log likelihood	-22.31279	Hannan-Quinn criter.	2.065591
F-statistic	20.14821	Durbin-Watson stat	1.687980
Prob(F-statistic)	0.000011		

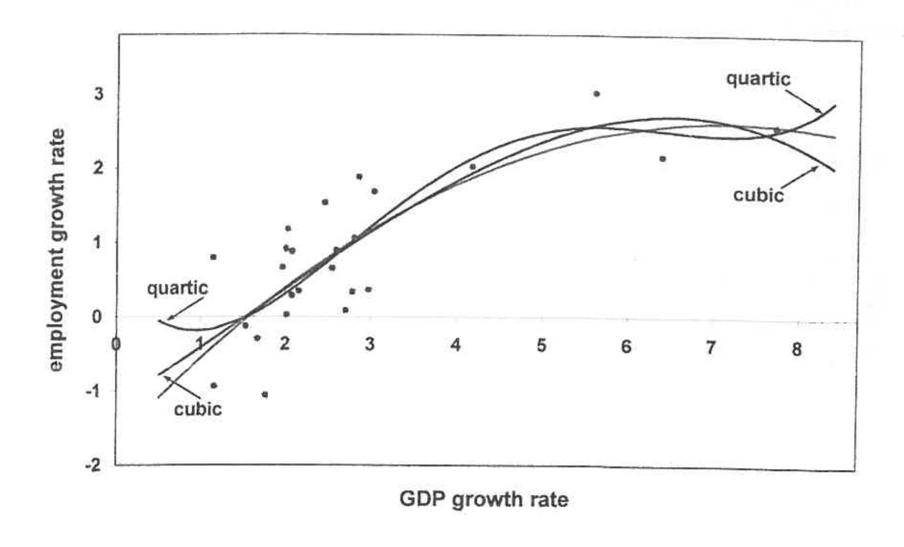


شكل (4 -11) انحدار قطع زائد ومربع معدل نمو العمالة على معدل نمو الناتج المحلي الاجمالي

4-3-2- كثير الحدود من مرتبة أعلى

لماذا نتوقف عند التربيع؟ لماذا لا نضيف التكعيب أو الدرجة الرابعة أو مرتبة أعلى؟ هناك عدة أسباب لعدم اضافتها منها أثر التناقص الحدي وهو معياري في النظرية الاقتصادية (الوصف التربيعي). لكن نادراً ما تقترح النظرية الاقتصادية علاقات تكعيبية أو أعلى. والسبب الثاني تحسين تقدير الحدود من رتبة أعلى، لكن هذه الحدود غير مسوغة نظرياً.

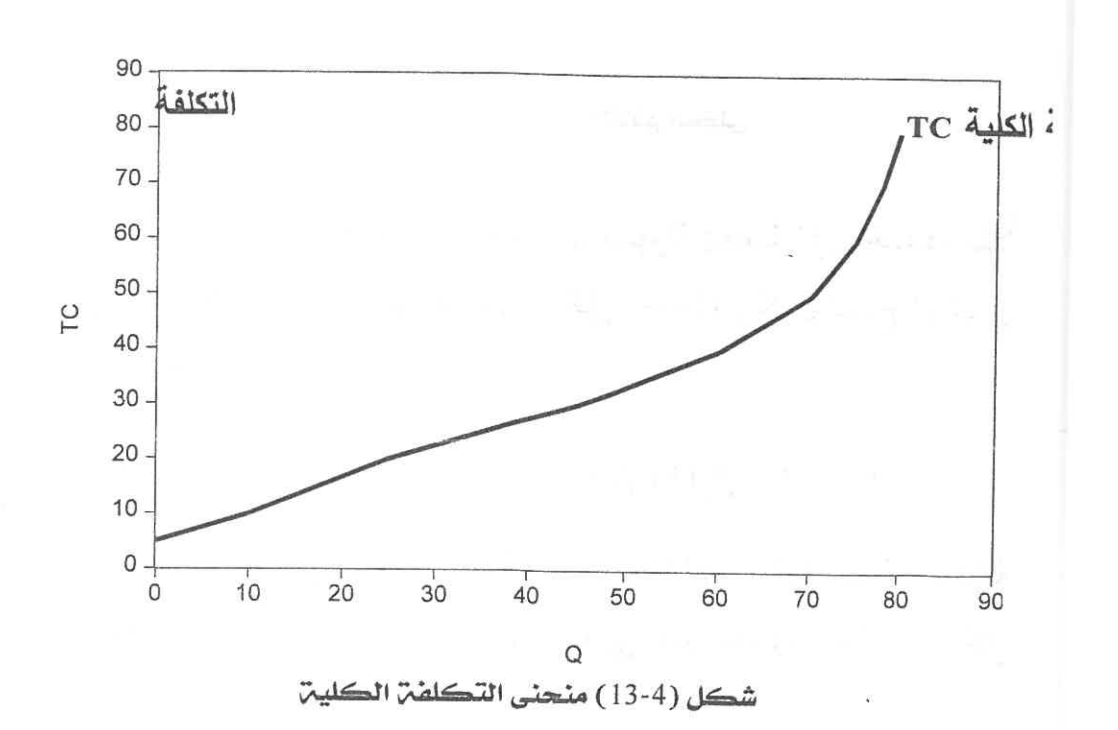
وهذه النقاط شرحها الشكل (4-12) الذي يظهر الانحدار التكعيبي والانحدار من الدرجة الرابعة مشابهين كثيراً للانحدار التربيعي.



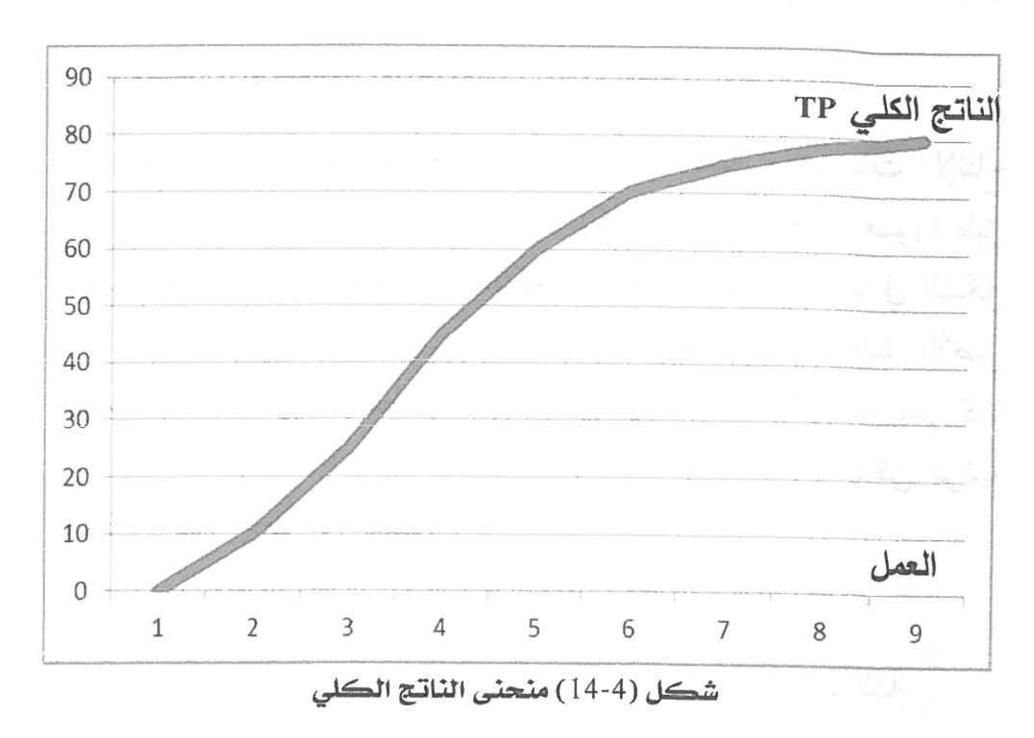
شكل (4-12) الانحدار التكعيبي ومن الدرجة الرابعة لانحدار معدل نمو العمالة على معدل نمو الناتج المحلي الاجمالي

تطبيق: منحنى التكاليف والإنتاج

ندرس في الاقتصاد الجزئي منحنيات التكلفة ومنحينات الإنتاج للمنشأة، ويعتبر كل من منحنى التكلفة الكلية والناتج الكلي صورة طبق الأصل عن الآخر، وتأخذ شكل التكعيب cubic المعياري كما في الشكل (4-13)، ومنحنيات التكلفة المتوسطة والتكلفة الحدية صورة طبق الأصل عن منحنيات الناتج المتوسط والناتج الحدي التي تأخذ الشكل التربيعي كما يعرضها الشكل (4-14)، وميل تلك العلاقات غير ثابت ولا يمكن عرضه بنموذج انحدار خطي في معلماته.



شكل (4-13) منحنى التكلفة الكلية



شكل (4-14) منحنى الناتج الكلي

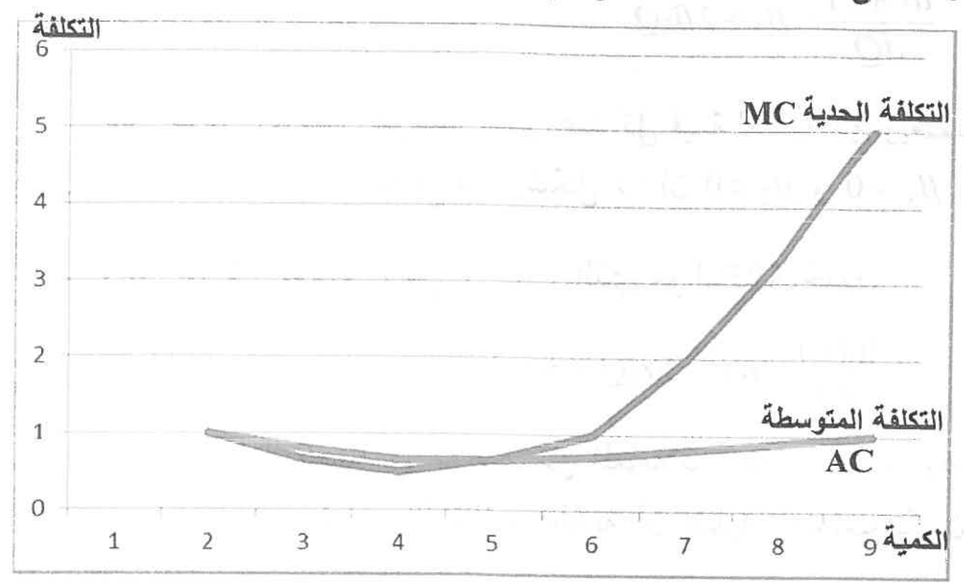
تستطيع عرض جميع تلك الأشكال بسهولة بمعادلة كثير الحدود؛ مثلاً نعبّر عن علاقة التكلفة المتوسطة في الشكل (4-14) ويكون نموذج الانحدار المناسب لها كما يلي:

$$AC = \beta_1 + \beta_2 Q + \beta_3 Q^2 + u \qquad (4.36)$$

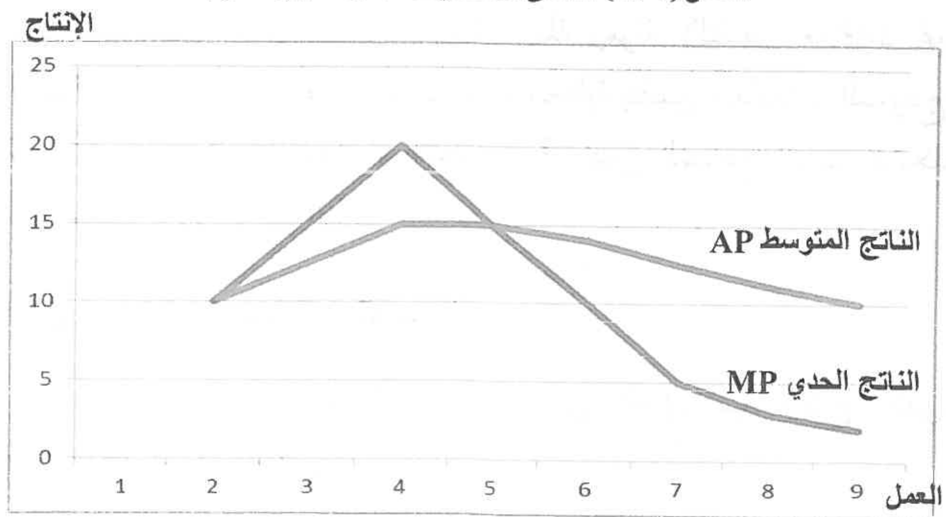
ويأخذ الشكل التربيعي شكل (∪) الذي يرتبط بدالة التكلفة المتوسطة، وكل منحنى تكاليف كلي تكعيبي كثير الحدود كما في الشكل (4-14) ويكون كما يلي:

$$TC = \alpha_1 + \alpha_2 Q + \alpha_3 Q^2 + \alpha_4 Q^3 + u \qquad (4.37)$$

تبيّن أشكال الدالة هذه أشكال غير خطية، ولا زلنا نستخدم طريقة المربعات الصغرى لتقديرها، والمتغيّران Q^2 و Q^3 هما متغيّران تفسيريان والتعامل معهما لا يختلف عن غيرهما.



شكل (4-15) منحنى التكاليف المتوسطة والحدية



شكل (4-16) منحنى الناتج المتوسط والناتج الحدي

الجانب الممتع في علاقات النماذج غير الخطية هو تفسير المعاملات لم تعد ميلاً، بالإضافة إلى ميل منحنى التكلفة المتوسطة للمعادلة (4.36) هو:

$$\frac{d(AC)}{dQ} = \beta_2 + 2\beta_3 Q \qquad (4.38)$$

ميل منحنى التكلفة المتوسطة يتغيّر عند كل قيمة للكمية Q ويعتمد على المعلمتين β_2 و β_3، ونتوقع من الشكل $\cup$ أن $\beta_2 < 0$ و $\beta_3 > 0$.

وميل منحنى التكلفة الكلي (4.37) الذي هو التكلفة الحدية:

$$\frac{d(TC)}{dQ} = \alpha_2 + 2\alpha_3 Q + 3\alpha_4 Q^2 \qquad (4.39)$$

وميل الدالة التكعيبية في Q يتضمّن المعاملات α_2 و α_3 و α_4 ونتوقع من شكل $\cup$ لمنحنى التكاليف الحدية أن إشارة المعلمات تكون $\alpha_2 > 0$ و $\alpha_3 < 0$ و $\alpha_4 > 0$.

استخدام متعدد الحدود طريقة سهلة ومرنة لالتقاط العلاقات غير الخطية بين المتغيّرات، إلا أنه يجب علينا العناية بتفسير معاملات النموذج، كما أن المتغيّرات التربيعية أو التكعيبية في نفس النموذج تسبب مشكلة الارتباط الخطي collineartiy.

مثال: معادلة التكلفة المتوسطة

سنقدر دالة التكاليف المتوسطة التي هي دالة في كمية الإنتاج الكلي، والتي تأخذ الشكل التالي:

$$AC = \beta_1 + \beta_2 Q + \beta_3 Q^2 + u$$

وللحصول على شكل ∪، سنتوقع أن $\beta_2 < 0$ و $\beta_3 > 0$، والأثر الحدي لأثر الكمية المنتجة على التكاليف المتوسطة يكون كما يلي:

$$\frac{d(AC)}{dQ} = \beta_2 + 2\beta_3 Q$$

ويكون أدنى حد للتكاليف- الكمية عندما تكون الكمية $Q = -\beta_2 / 2\beta_3$، وعند هذه النقطة يكون الميل صفراً.

جدول (4-2) الإنتاج والتكاليف الكلية والحدية والمتوسطة

عدد العمال L	الإنتاج TP Q	الناتج المتوسط AP	الناتج الحدي MP	السعر P	الأجور W	الإيراد الكلي TR	الإيراد الحدي MR	كلفة العمال TLC	الكلفة المتوسطة AC	كلفة العمل الحدي MLC
0	0	0	0	2	10	0		0		0
1	10	10	10	2	10	20	20	10	1.0	10
2	25	12.5	15	2	10	50	30	20	0.8	10
3	45	15	20	2	10	90	40	30	0.7	10
4	60	15	15	2	10	120	30	40	0.7	10
5	70	14	10	2	10	140	20	50	0.7	10
6	**75**	**12.5**	**5**	**2**	**10**	**150**	**10**	**60**	**0.8**	**10**
7	78	11.143	3	2	10	156	6	70	0.9	10
8	80	10	2	2	10	160	4	80	1.0	10

استخدمنا بيانات الجدول (4-2) لتقدير معادلة التكاليف المتوسطة، وكانت نتائج معاملات الكمية ومربع الكمية كما هو متوقع لإشاراتها ومعاملاتها معنوية إحصائياً، وكانت النتائج كما يلي:

Dependent Variable: AC
Method: Least Squares
Date: 12/18/15 Time: 14:46
Sample (adjusted): 2002 2009
Included observations: 8 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.256889	0.076592	16.41019	0.0000
Q	-0.026260	0.003965	-6.623366	0.0012
Q^2	0.000277	4.18E-05	6.612239	0.0012

R-squared	0.898346	Mean dependent var	0.818132
Adjusted R-squared	0.857684	S.D. dependent var	0.136154
S.E. of regression	0.051364	Akaike info criterion	-2.819762
Sum squared resid	0.013191	Schwarz criterion	-2.789971
Log likelihood	14.27905	Hannan-Quinn criter.	-3.020687
F-statistic	22.09311	Durbin-Watson stat	1.050397
Prob(F-statistic)	0.003295		

$$AC = 1.2568 - 0.02626\ Q + 0.000276\ Q^2$$

ولتقدير الأثر الحدي للكمية 60 نحصل على ما يلي:

$$\frac{d(\hat{A}C)}{dQ} = - 0.02626 + 2(\ 0.000276)\ (60) = -0.0252$$

قدرنا عند الكمية 60 وأن زيادة كمية أضافية ستنخفض التكلفة بمقدار 2.52 وحدة تكلفة، ونقطة الانعطاف في العلاقة بعد أي كمية ستبدأ التكاليف بالتزايد، وسنقدر حدوثه عند الكمية:

$$Q = -\beta_2 / 2\beta_3$$

$$= - \frac{- 0.02626}{2 \times (\ 0.000276)}$$

$$= 47.57$$

4-3-3- المتغيّرات التفسيرية التفاعلية

ثم ننتقل إلى نماذج بحدود تفاعلية مثل:

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_2 X_3 + u \qquad (4.40)$$

هذا نموذج خطي في معلماته ويقدر باستخدام المربعات الصغرى، وفي الحقيقة هو غير خطي في متغيّراته التفسيرية ومعلماته معقدة، وليس من الممكن تفسير β_2 كأثر للمتغيّر X_2 على Y مع بقاء X_3 و X_2X_3 ثابتة؛ لأنه ليس من الممكن إبقاء X_3 و X_2X_3 ثابتة إذا تغيّر X_2.

نستطيع اعادة كتابة النموذج لاعطاء تفسير مناسب للمعلمات كما يلي:

$$Y = \beta_1 + (\beta_2 + \beta_4 X_3) X_2 + \beta_3 X_3 + u \qquad (4.41)$$

هذا يُبيّن حقيقة ضمنية هي أن $(\beta_2 + \beta_4 X_3)$ الأثر الحدي للمتغيّر X_2 على Y يعتمد على قيمة X_3، ومن هذا نستطيع أن نرى أن تفسير المعامل β_2 له تفسيراً خاصاً، وهو يعطي الأثر الحدي للمتغيّر X_2 على Y عندما $X_3 = 0$.

أو نعيد كتابة النموذج كما يلي:

$$Y = \beta_1 + \beta_2 X_2 + (\beta_3 + \beta_4 X_2) X_3 + u \qquad (4.42)$$

ونرى من هذا أن الأثر الحدي للمتغيّر X_3 على Y مع بقاء X_2 ثابتاً يكون $(\beta_3 + \beta_4 X_2)$ وقد تفسر المعلمة β_3 كأثر حدي للمتغيّر X_3 على Y عندما $X_2 = 0$.

إذا كان $X_3 = 0$ خارج نطاق X_3 في العينة، يتم تفسير تقدير β_2 كتقدير للأثر الحدي للمتغيّر X_2 عندما تكون $X_3 = 0$ ويجب أن تعالج بحذر، في بعض الأوقات يكون التقدير بنفس الطريقة مستحيلاً؛ كتقدير الحد الثابت في الانحدار مستحيل إذا أُعطيت تفسيراً حرفياً، وسنواجه نفس المشكلة بتفسير β_2 في التوصيف التربيعي، ومن الممتع مقارنة تقدير أثر X_2 و X_3 على Y في نموذج لا يتضمن حد تفاعلي، والتغيّر في معنى β_2 و β_3 بسبب تضمين حد تفاعلي يجعل هذا التفاعل صعباً.

مثال: التفاعل بين المتغيّرات المستمرة

إذا اشتمل نموذج الانحدار على متغيّرين مستمرين سيكون الأثر لتغيّر العلاقة بين المتغيّرين كل منهما على الآخر وعلى المتغيّر التابع، مثلاً سنأخذ نموذج دورة الحياة لشرح هذه الفكرة:

افرض أننا نرغب بدراسة أثر الدخل والعمر على إنفاق الأفراد على البيتزا، ولتحقيق هذا الهدف سنأخذ عينة عشوائية مكونة من 45 شخصاً أعمارهم 18 سنة فأكبر، وسجلنا نفقاتهم السنوية على البيتزا (Pizza)، إضافة إلى دخلهم (Income) والعمر (Age)، كما تظهر البيانات كاملة في الجدول (4-2).

سيكون النموذج الأولي كما يلي:

$$Pizza = \beta_1 + \beta_2\ Age + \beta_3\ Income + u \qquad (4.43)$$

جدول (4-2) الطلب على البيتزا

PIZ	Income	AGE
109	15000	25
0	30000	45
0	12000	20
108	20000	28
220	15000	25
189	30000	35
64	12000	40
262	12000	22
64	28000	30
35	22000	21
94	44000	40
71	10000	21
403	22200	0 45
41	32000	36
10	45000	36
110	55000	40
239	29000	23
63	39000	32
0	70000	52
106	55000	30
0	90000	45
141	6000	32
299	18000	20
148	55000	55
424	10000	18
242	23000	30
119	35000	45
338	38000	40
135	45000	50
590	85000	32
324	22000	30
87	25000	51
395	29000	22
513	13200	0 40
56	35000	30
400	80000	36
384	55000	27
262	30000	24
336	27000	21
281	80000	45

ستكون آثار هذا الوصف كما يلي:

1- $\partial(Pizza)/\partial Age = \beta_2$: عند مستوى الدخل المحدد، يُتوقع أن يتغيّر الإنفاق على البيتزا بمقدار β_2 لكل سنة إضافية للعمر، ونتوقع أن تكون إشارة β_2 سالبة، أي سينخفض الانفاق على البيتزاء مع زيادة العمر، بدون أثر الدخل.

2- $\partial(Pizza)/\partial Income= \beta_3$: عند العمر المحدد، عندما يزيد الدخل 1 دولار ستتوقع أن يكون الإنفاق على البيتزاء بمقدار β_3 ، وبما أن البيتزا سلعة عادية normal good ستتوقع أن تكون الإشارة موجبة، وتسمى المعلمة β_3 بالميل الحدي للإنفاق على البيتزا.

وتم تقدير المعادلة (4.43) كما يلي: (قيمة t بين الأقواس)

$$\underset{(t)}{Pizza} = 342.88 - \underset{(-3.27)}{7.58}\ Age + \underset{(3.95)}{0.0024}\ Income$$

وكانت إشارة المعلمات المقدّرة كما هو متوقع، ومعلمات كل من Age و Income معنوية اعتماداً على إحصائية t الخاصة لكل معلمة.

ومن المنطقي أن نتوقع بغض النظر عن عمر الشخص، أن زيادة الدخل بمقدار 1 دولار ستؤدي إلى زيادة الإنفاق على البيتزا بمقدار β_3 دولار أم لا؟ قد يكون لا، ويبدو من المعقول أن نفترض أن الشخص يزداد عمره وينخفض ميلة للإنفاق على البيتزا، وفي هذه الحالة يعتمد أثر الدخل

على عمر الشخص، وبالتالي سيُعدّل أثر أحد المتغيّرات بالآخر، ولحساب مثل هذه التفاعلات يتم تضمين متغيّر تفاعلي interaction variable يكون بحاصل ضرب متغيّرين من متغيّرات المعادلة، وبما أن المتغيرين *Age* و *Income* هما المتغيّران اللذان سيتفاعلان سنضيف متغيّراً جديداً $(Age \times Income)$ إلى نموذج الانحدار، وتكون النتيجة كما يلي:

$$Pizza = \beta_1 + \beta_2\ Age + \beta_3\ Income + \beta_4 (Age \times Income) + u \quad (4.44)$$

عندما يتم تضمين النموذج بمتغيّرين مستمرين سيتطلب منا الحذر عند تفسير المعلمات، ويكون أثر *Age* و *Income* كما يلي:

1- $\partial(Pizza)/\partial Age = \beta_2 + \beta_4\ Income$: يعتمد أثر العمر *Age* الآن على الدخل المحدد، وحسب عمر الشخص يُتوقع انخفاض الإنفاق على البيتزا، ويتوقع أن تكون إشارة β_4 سالبة؛ أي سينخفض الانفاق على البيتزا مع زيادة العمر، وبزيادة الدخل ستنخفض الزيادة بتغيّر العمر.

2- $\partial(Pizza)/\partial Income = \beta_3 + \beta_4\ Age$: أثر تغيّر الدخل على الإنفاق على البيتزا هو الميل الحدي للإنفاق على البيتزا، ويعتمد على *Age*، فإذا كان منطقنا صحيحاً ستكون β_4 سالبة، وإذا زاد *Age* ستنخفض قيمة المشتقة الجزئية.

يتضمن النموذج المقدّر للمعادلة (4.44) ناتج $(Age \times Income)$ كما يلي:

$$\underset{(t)}{Pizza} = 161.47 - \underset{(-0.89)}{2.98}\, Age + \underset{(2.47)}{0.009}\, Income - \underset{(-1.85)}{0.00016}\,(Age \times Income)$$

المعامل المقدّر للحد التفاعلي سالب ومعنوي عند مستوى معنوية $\alpha = 0.05$ ، وبقيت إشارة المعاملات كما هي، إلا أن معامل Age غير معنوي، وهذا يُبيّن أن Age يؤثر على الانفاق على $Pizza$ من خلال تفاعلة مع الدخل فقط، وهو الميل الحدي للإنفاق على البيتزا.

وباستخدام هذا التقدير لتقدير الأثر الحدي للعمر على الانفاق على البيتزا لشخصين: دخل أحدهما 25000 دولار، ودخل الثاني 90000 دولار.

$$\frac{\partial(\hat{P}izza)}{\partial Age} = b_2 + b_4\, Income$$

$$= -2.98 - 0.00016\, Income$$

$$= \begin{cases} -6.98 & for\, Income = \$25{,}000 \\ -17.40 & for\, Income = \$90{,}000 \end{cases}$$

نتوقع أن الشخص الذي دخله 25000 دولار سيُخفّض الانفاق على البيتزا بمقدار 6.98 دولار في السنة، بينما الشخص الذي دخله 90000 دولار سيخفّض الإنفاق على البيتزا بمقدار 17.40 دولار مع بقاء جميع العوامل الأخرى ثابتة.

كما أن إحدى طرق تخفيف المشكلة هي اعادة قياس X_2 و X_3 بقياس الانحراف عن وسط العينة:

$$X_2^* = X_2 - \overline{X}_2 \qquad (4.45)$$

$$X_3^* = X_3 - \overline{X}_3 \qquad (4.46)$$

ونعوض x_2 و X_3 ويصبح النموذج كما يلي:

$$Y = \beta_1 + \beta_2\left(X_2^* + \overline{X}_2\right) + \beta_3\left(X_3^* + \overline{X}_3\right) + \beta_4\left(X_2^* + \overline{X}_2\right)\left(X_3^* + \overline{X}_3\right) + u$$
$$= \beta_1^* + \beta_2^* X_2^* + \beta_3^* X_3^* + \beta_4 X_2^* X_3^* + u \qquad (4.47)$$

حيث أن $\beta_1^* = \beta_1 + \beta_2\overline{X}_2 + \beta_3\overline{X}_3 + \beta_4\overline{X}_2\overline{X}_3$ و $\beta_2^* = \beta_2 + \beta_4\overline{X}_3$ و $\beta_3^* = \beta_3 + \beta_4\overline{X}_2$، وما قمنا به هو أن معامل X_2^* و X_3^* يعطي الأثر الحدي للمتغيّرات مع بقاء المتغيّر الآخر عند متوسط العينة. مثلاً اعد كتابة المعادلة الجديدة:

$$Y = \beta_1^* + \left(\beta_2^* + \beta_4 X_3^*\right)X_2^* + \beta_3^* X_3^* + u \qquad (4.48)$$

كما هو ظاهراً، يعطي β_2^* الأثر الحدي للمتغيّر X_2^* وبالتالي X_2 عندما $X_3^* = 0$، و X_3 عند متوسط عينتها، وتفسر β_3^* كما سبق.

4-3-4- اختبار رامزي Ramsey's RESET لسوء تحديد النموذج

للكشف عن امكانية أن يكون المتغيّر التابع في النموذج دالة غير خطية يتم اضافة الحد التربيعي للمتغيّرات التفسيرية والحدود التفاعلية لتحديد النموذج، فإذا كان في النموذج عدة متغيّرات تفسيرية، باستخدام اختبار Ramsey's RESET test للكشف عن سوء تحديد الدالة، يزودنا

بمؤشر بسيط. ولتطبيقه ننفذ أولاً الانحدار بشكله الأصلي ونخزن القيم المقدرة للمتغيّر التابع المشار إليها $\hat{Y}$ وهي كما يلي:

$$\hat{Y} = b_1 + \sum_{j=2}^{k} b_j X_j \qquad (4.49)$$

يعتبر $\hat{Y}^2$ مزيج خطي لمربع المتغيّر X وتفاعلاته؛ فإذا تم اضافة $\hat{Y}^2$ إلى محددات الانحدار سيلتقط التربيع والتفاعل غير الخطي، فإذا لم يكن ضرورياً فإنه يزيد ارتباطه مع أي متغيّر مستقل X، ويستهلك درجة حرية واحدة. ويتم اختبار معلمة $\hat{Y}^2$ فإذا كانت احصائية t لمعلمته معنوية، فهذا يشير إلى وجود علاقة غير خطية.

بالطبع، فإن هذا الاختبار لا يستطيع تحديد الشكل الحقيقي غير الخطي، وقد يفشل عن كشف الأنواع غير الخطية الأخرى. ومن حيث المبدأ قد يضاف $\hat{Y}$ بقوة (أس) أعلى، وبالتالي يظهر أنه غير جدير بالاهتمام؛ وإذا تم إضافة حدود بأس أعلى مثل $\hat{Y}^3$ و $\hat{Y}^4$ يتم استخدام اختبار F.

4-4- النماذج غير الخطية

افرض أن المتغيّر Y يعتمد على المتغيّر X حسب العلاقة التالية:

$$Y = \beta_1 + \beta_2 X^{\beta_3} + u \qquad (4.50)$$

وترغب بالحصول على تقدير β_1 و β_2 و β_3 من بيانات Y و X، ولا يوجد أي طريقة لتحويل (4.50) للحصول على علاقة خطية. وبالتالي، من غير الممكن تطبيق إجراءات الانحدار العادي.

ومع ذلك لا زالت امكانية استخدام مبدأ تخفيض مجموع مربع البواقي للحصول على تقدير المعلمات. سيتم وصف الانحدار **اللوغاريتمي غير الخطي البسيط nonlinear regression algorithm** كما يلي:

1- إبدأ بالقيمة المحتملة المعقولة للمعلمات.

2- احسب القيم المتوقعة للمتغيّر التابع Y من بيانات X باستخدام قيم تلك المعلمات.

3- احسب بواقي مشاهدات العينة ثم مجموع مربع البواقي SSR.

4- اعمل تغيير بسيط في أحد المعلمات المقدّرة أو أكثر.

5- احسب قيم Y المتوقعة الجديدة وبواقيها و SSR.

6- إذا كان SSR أقل مما كانت عليه من قبل؛ فإن تقدير المعلمات الجديد يكون أفضل من السابق واعتبره نقطة بداية جديدة.

7- أعد الخطوة 4 و 5 و 6 مرة أخرى إلى أن تصل إلى قيم غير قابلة للتغيير في تقدير المعلمات التي تخفض SSR.

8- نستنتج أنه لديك أقل SSR وتصف تقدير المعلمات النهائي كتقدير المربعات الصغرى.

تمارين

4-3- تُظهر نتائج انحدار Y على X و X^2 أدناه.

والمطلوب شرح سبب الإشارة السالبة لمعامل X.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
X	-0.2564658	0.1318583	-1.95	0.052
X^2	0.0271172	0.0060632	4.47	0.000
c	12.79121	0.7366358	17.36	0.000

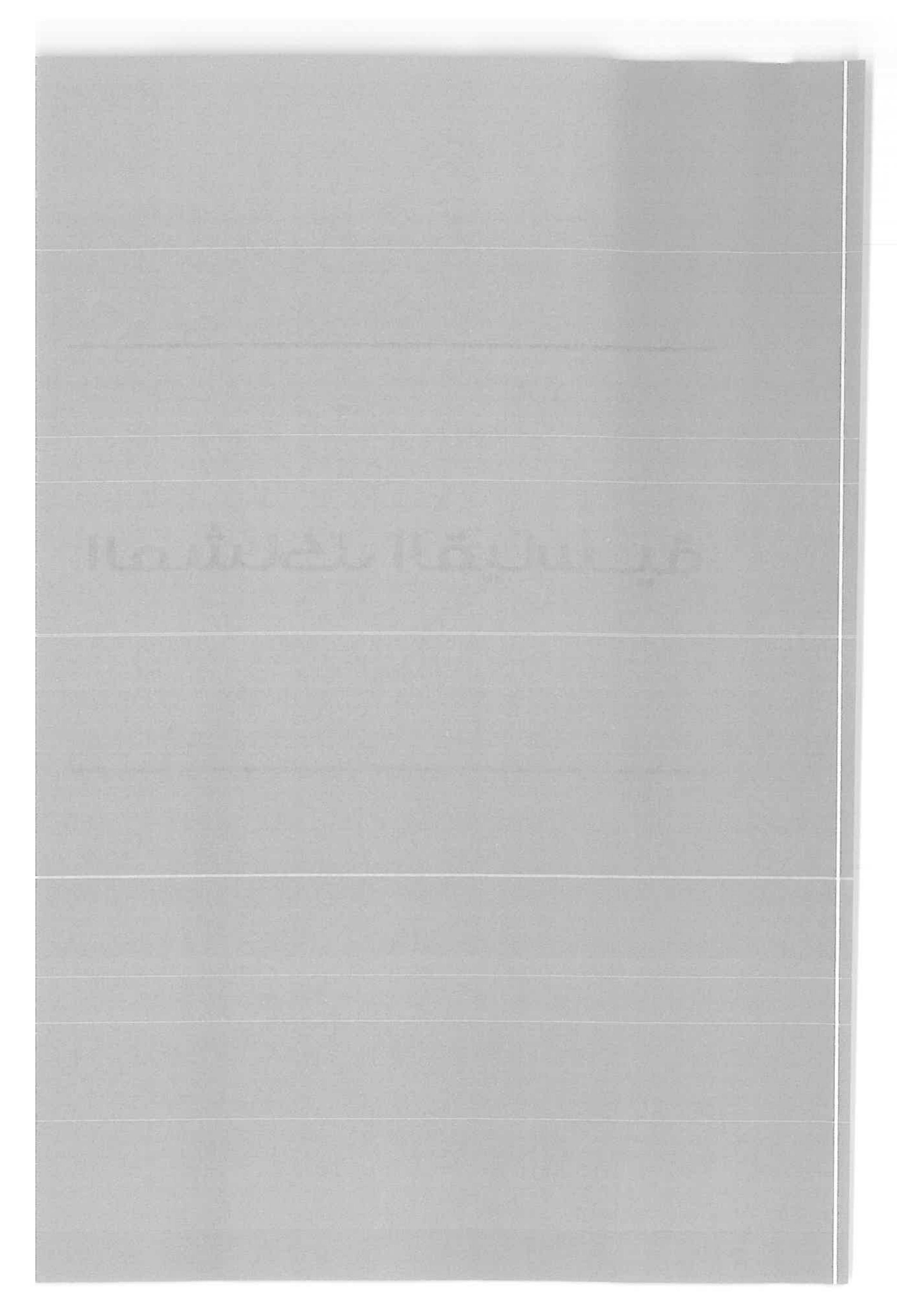
4-4- يظهر الجدول أعلاه نتائج انحدار LnY على LnX و LnX^2.

المطلوب شرح نتائج الانحدار.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LnX	omitted			
LnX^2	0.100341	0.0132915	7.55	0.000
c	2.113730	0.0648636	32.59	0.000

4-5- نفذ اختبار RESET لسوء توصيف الشكل الدالي باستخدام بيانات الاستهلاك العام G على GDP، وخزن قيم التقدير باسم Yhat مثلاً، وعرّف Yhatsq كمربع Yhat، واضف Yhatsq إلى توصيف المعادلة واختبرمعاملاتها.

المشاكل القياسية



الفصل الخامس

الارتباط الخطي المتعدد

Multicollinearity

تبحث الفصول الثلاثة التالية بانتهاك الفرضيات التقليدية وعلاجها، ويبحث هذا الفصل الارتباط الخطي المتعدد والفصلين التاليين يبحثان عدم تجانس التباين والارتباط الخطي المتسلسل، وسنحاول الإجابة عن الأسئلة التالية:

1- ما هي طبيعة المشكلة.

2- ما هي نتائج هذه المشكلة؟

3- كيف نشخص هذه المشكلة؟

4- ما هو علاج هذه المشكلة؟

عندما نقول الارتباط الخطي المتعدد يعني الحديث عن انتهاك للفرضية (8) التي تنص على ألاَّ يكون أي متغيّر مستقل دالة خطية تامة في أحد المتغيّرات المستقلة أو اكثر، وهذه الحالة نادرة الحدوث، إلا أن قوة الارتباط الخطي غير التام لا تنتهك الفرضية (8) وتسبب مشكلة دائمة.

يخبرنا معامل المتغيّر β_k عن تأثيره على المتغيّر التابع بزيادة المتغيّر المستقل X_k بوحدة واحدة مع بقاء المتغيّرات المستقلة الأخرى في المعادلة ثابتة، إلا إذا كان بين المتغيّرين التفسيريين ارتباط قوي في العينة، فعندما يتغيّر أحدهما سيميل الآخر للتغيّر كذلك، وسنجد صعوبة في تمييز أثر أحد المتغيّرات على الآخر، وبما أن المتغيّرات X_s تستطيع التحرك مع بعضها في العينة، فإن خطورة الارتباط الخطي المتعدد قد تختلف كثيراً.

فإذا كان ارتباطاً قوياً جداً بين متغيّرين مستقلين أو أكثر يكون من الصعب الحصول على تقدير دقيق لمعلمات نموذجٍ صحيح، وإذا كان تحرك المتغيّرين متماثل فلا يوجد أمل للتمييز بين أثرهما، إلا إذا كانت المتغيّرات مرتبطة فقط، فإننا لا نزال نستطيع تقدير أثرهما بدقة كافية لأغلب الأهداف.

5-1- الارتباط الخطي المتعدد التام وغير التام

5-1-1- الارتباط الخطي المتعدد التام Perfect Mulicollinearity

لفهم الارتباط الخطي المتعدد التام نأخذ النموذج التالي:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \qquad (5.1)$$

حيث أن القيم المفترضة للمتغيّرين المستقلين X_2 و X_3 هي:

X_2	1	2	3	4	5	6
X_3	2	4	6	8	10	12

ونلاحظ أن $X_3 = 2X_2$، وبالتالي فإن المعادلة (5.1) تحتوي على متغيّرين تفسيريين مستقلين X_2 و X_3، إلا أن المعلومات توضح أن المتغير

X_3 ليس مستقلاً عن المتغيّر X_2، لأن X_3 هو دالة خطية في X_2، ونقول في هذه الحالة أن X_2 و X_3 هما مرتبطان خطياً، وهذا يعني وجود ارتباط خطي تام[1] بينهما إذا كان أحد المتغيّرين دالة خطية بالمتغيّر الآخر، وعندما يحدث هذا تكون المعادلة كما يلي:

$$\delta_1 X_2 + \delta_2 X_3 = 0 \qquad (5.2)$$

وتتمتع δ_1 و δ_2 بقيم غير صفرية. ولدينا في هذا المثال $X_3 = 2X_2$، وعليه $(-2)X_2 + (1)X_3 = 0$؛ أي أن $\delta_1 = -2$ و $\delta_2 = 1$، ومن الواضح أنه عندما يكون الحل في (5.1) يساوي 0 $\delta_1 = \delta_2 = 0$ (يسمى حل عديم الأهمية) ويكون X_2 و X_3 مستقلين خطياً، ويتطلب غياب الارتباط الخطي التام عدم وجود حالة (5.2) بالضبط.

وفي حالة أكثر من متغيّر تفسيري (5 متغيّرات مثلاً)، ستكون حالة الارتباط الخطي عندما يكون أحد المتغيّرات دالة خطية تامة بأحد المتغيّرات الأخرى، أو أكثر من متغيّر، أو كلها. وتكون الصيغة كما يلي:

$$\delta_1 X_1 + \delta_2 X_2 + \delta_3 X_3 + \delta_4 X_4 + \delta_5 X_5 = 0 \qquad (5.3)$$

يكون على الأقل معاملين اثنين غير صفريين.

ولفهم أفضل لهذه الحالة، ما تبينه مصيدة المتغيّرات الصورية أو الوهمية Dummy variable trap، افرض أن X_1 المقطع أو الحد الثابت

[1] تعني كلمة تام (Perfect) أننا نستطيع شرح تغيّر أحد المتغيّرات المستقلة من خلال تحركات متغيّر تفسيري آخر؛ أي أن لهما نفس الأثر.

$(X_1 = 1)$ على سبيل المثال، وتكون X_2 و X_3 و X_4 و X_5 فصولاً وهمية لبيانات سلاسل زمنية ربع سنوية (مثلاً تأخذ X_2 القيمة 1 للفصل الأول وصفر لما عداها، و X_3 القيمة 1 للفصل الثاني وصفر لما عداها، وهكذا ...)، وبالتالي يكون لدينا في هذه الحالة $X_2 + X_3 + X_4 + X_5 = 1$، ولأن $X_1 = 1$ فإن $X_1 = X_2 + X_3 + X_4 + X_5$، ويكون الحل $\delta_1 = 1$، و $\delta_2 = -1$، و $\delta_3 = -1$، و $\delta_4 = -1$، و $\delta_5 = -1$ وهذه المجموعة من المتغيّرات مرتبطة خطياً.

سترى في حالة الارتباط الخطي التام أن تقدير OLS ليس فريداً، ولمزيد من التوضيح خذ على سبيل المثال النموذج التالي:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \qquad (5.4)$$

حيث أن $X_3 = \delta_1 + \delta_2 X_2$، وتسمى δ_1 و δ_2 قيمتان ثابتتان، ثم عوّض هذا في (5.4):

$$\begin{aligned} Y_i &= \beta_1 + \beta_2 X_{2i} + \beta_3(\delta_1 + \delta_2 X_2) + u_i \\ &= (\beta_1 + \beta_3\delta_1) + (\beta_2 + \beta_3\delta_2) X_2 + u_i \\ &= \quad \upsilon_1 \quad + \quad \upsilon_2 \quad X_2 + \varepsilon \qquad (5.5) \end{aligned}$$

حيث أن $\upsilon_1 = \beta_1 + \beta_3\delta_1$ و $\upsilon_2 = \beta_2 + \beta_3\delta_2$.

سيتم تقدير المعاملين υ_1 و υ_2 بغض النظر عن جودتهما، ولا نحصل على تقدير فريد للمعلمات β_1 و β_2 و β_3، وللحصول عليها علينا حل المعادلتين التاليتين:

$$\hat{\upsilon}_1 = \hat{\beta}_1 + \hat{\beta}_3 \delta_1$$
$$\hat{\upsilon}_2 = \hat{\beta}_2 + \hat{\beta}_3 \delta_2$$

هذا نظام معادلتين بثلاث مجاهيل هي $\hat{\beta}_1$ و $\hat{\beta}_2$ و $\hat{\beta}_3$، وفي أي نظام يحوي متغيّرات أكثر من المعادلات، سيحتوي على عدد لا نهائي من الحلول. مثلاً اختر قيمة اعتباطية لقيمة $\hat{\beta}_3$ ولتكن k، فإن $\hat{\beta}_3 = k$ نجد $\hat{\beta}_1$ و $\hat{\beta}_2$ كما يلي:

$$\hat{\beta}_1 = \hat{\upsilon}_1 - \delta_1 k$$
$$\hat{\beta}_2 = \hat{\upsilon}_2 - \delta_2 k$$

وبما أنه يوجد قيم لا نهائية يتم استخدامها لـ k يكون لدينا عدد لا نهائي من حلول $\hat{\beta}_1$ و $\hat{\beta}_2$ و $\hat{\beta}_3$. وإذا وجد الارتباط الخطي التام فإنه لا يوجد أي منهجية تقدير تستطيع تزويدنا بتقدير فريد لمعلمات المجتمع، ففي صيغة المصفوفات، أو لحالة أكثر عمومية، إذا كان أحد أعمدة المصفوفة X دالة خطية تامة بأحد أو اكثر من الأعمدة الأخرى تكون المصفوفة $X'X$ فريدة *Singular*، وهذا يعني أن محددها سيكون صفراً $(|X'X| = 0)$، علماً بأن مُقدّرات OLS تحسب كما يلي:

$$\hat{\beta} = (X'X)^{-1} X'Y$$

ويتم عكس المصفوفة (أخذ المعكوس أو النظير الضربي) كما يلي:

$$(X'X)^{-1} = \frac{1}{|X'X|} adj.(X'X)$$

ولأن $|X'X| = 0$ لا نستطيع عكسها.

نادراً ما يحدث الارتباط الخطي التام ببيانات فعلية، إلا أنه يحدث من الأخطاء التصحيحية مثل مصيدة المتغيّرات الوهمية، أو تضمين المعادلة متغيّرات مثل $Ln X$ و $Ln X^2$ بنفس الوقت.

5-1-2- الارتباط الخطي غير التام Imperfect Mulicollinearity

يحدث الارتباط الخطي غير التام عندما تكون المتغيّرات التفسيرية في المعادلة مرتبطة ارتباطاً غير تام، ويمكن التعبير عن الارتباط الخطي غير التام كما يلي: عندما تكون العلاقة بين المتغيّرات المستقلة في (5.4) مثل $X_3 = X_2 + \upsilon$، حيث υ المتغيّر العشوائي الذي يمثل "الخطأ" في العلاقة التامة بين المتغيّرين، تكون υ قيمة غير صفرية، لذا نستطيع الحصول على تقدير OLS. وفي الواقع، كل معادلة انحدار متعدد تحتوي ارتباطاً بين متغيّراتها التفسيرية. مثلاً تحتوي بيانات السلاسل الزمنية اتجاهاً عاماً صاعداً يسبب ارتباط عالٍ للمتغيّرات، وبالتالي فإن ارتفاع درجة الارتباط الخطي المتعدد في إحدى العلاقات يكون كافياً لخلق مشكلة، وقبل أن نترك هذه النقطة نحتاج إلى اختبار أثر الارتباط الخطي غير التام في مُقدّرات OLS.

5-2- مشاكل الارتباط الخطي المتعدد

إذا وجد الارتباط الخطي المتعدد في عينة، ماذا سيحدث للتقدير المحسوب من العينة؟ وبما أن الارتباط الخطي المتعدد التام يعني أن تقدير المعادلة غير ممكن، ماذا ستعني نتيجة الارتباط الخطي المتعدد غير التام؟ الهدف من هذا الجزء شرح نتائج الارتباط الخطي المتعدد واستكشاف بعض الأمثلة لهذه النتائج.

5-2-1- ما هي نتائج الارتباط الخطي المتعدد

النتائج الرئيسية للارتباط الخطي المتعدد:

1- يكون التقدير غير متحيّز.

2- زيادة التباين والخطأ المعياري للتقدير.

3- انخفاض إحصائية t المحسوبة.

4- يصبح التقدير حساساً لتغيير وصف المعادلة؛ حيث أن إضافة أو حذف المتغيّرات يسبب تغيّرات في قيم $\hat{\beta}$ عندما يوجد ارتباط خطي متعدد.

5- تقدير المعادلة الكاملة وتقدير معاملات متغيّر تكون غير مؤثرة: تكون إحصائية t منخفضة في معادلة الارتباط الخطي المتعدد، ويكون $\overline{R}^2$ مرتفعاً لتقدير المعادلة الكاملة، وبالتالي تكون معاملات الانحدار الفردية غير معنوية.

مثال: نتيجة الارتباط الخطي المتعدد

لنرى هل نستطيع تقدير معادلة تحوي ارتباطاً خطياً متعدداً قوياً، دعنا نرى مثالاً افتراضياً، افرض أننا قررنا تقدير دالة استهلاك للطلبة ونريد تقدير المعادلة التالية:

$$\hat{Cons} = f(\overset{+}{Yd}, \overset{+}{LA}) + \varepsilon = \beta_0 + \beta_1 YD + \beta_2 LA + \varepsilon \qquad (5.6)$$

حيث أن:

Cons: نفقات الاستهلاك السنوي للطلبة

Yd: الدخل السنوي المتاح للطلبة

LA: الأصول السائلة (مثل المدخرات، ..) للطلبة

ε : حد الخطأ العشوائي

تم جمع البيانات من الأفراد الذين يجلسون بجانبك في الفصل الدراسي، وكانت على النحو التالي:

LA	Yd	Cons	الطالب
25000	2500	2000	بُشرى
31000	3000	2300	مها
33000	3500	2800	عبدالله
39000	4000	3800	أحمد
48000	4500	3500	نور الهدى
54000	5000	5000	إيناس
55000	5500	4500	رائد

إذا قدرت انحداراً بطريقة المربعات الصغرى العادية على مجموعة بيانات على المعادلة (5.6) نحصل على:

$$\hat{Cons} = -367.83 + \underset{\substack{(1.0307)\\ t=0.496}}{0.5113}\, YD + \underset{\substack{(0.0942)\\ 0.453}}{0.0427}\, LA$$

$$\bar{R}^2 = 0.835 \qquad (5.7)$$

ومن جهة أخرى، إذا قدرنا دالة الاستهلاك كدالة في الدخل المتاح فقط سنحصل على:

$$\hat{Cons} = -471.43 + 0.9714\, YD$$
$$(0.157)$$
$$t = 6.187$$
$$\overline{R}^2 = 0.861 \qquad (5.8)$$

لاحظ من المعادلة (5.6) و (5.7) أن إحصائية t للدخل المتاح زادت بأكثر من 10 أضعاف عند اسقاط متغيّر الأصول السائلة من المعادلة، لماذ حدث هذا؟ أولاً أن معامل الارتباط البسيط بين YD و LA كان مرتفعاً $r_{YD,LA} = 0.986$، وهذه الدرجة المرتفعة من الارتباط تجعل الخطأ المعياري للمعامل المحسوب كبيراً جداً عندما تتضمّن المعادلة المتغيّرين معاً، وفي حالة $\hat{\beta}_{YD}$ ارتفع الخطأ المعياري من 0.157 إلى 1.03 عند إضافة LA إلى المعادلة، والمعامل المُقدّر نفسه تغيّر، بالإضافة إلى ذلك، أن $\overline{R}^2$ للمعادلتين متشابه بالرغم من الفرق الكبير في معنوية المتغيّرات التفسيرية في المعادلتين، وهذه النتائج هي نمط معادلات تحتوي ارتباط خطي متعدد.

أي معادلة هي الأفضل؟ إذا كان متغيّر الأصول السائلة نظرياً جزء من المعادلة، سيسبب اسقاطه خطورة تحيّز حذف أو اسقاط متغيّر، لكن وجود المتغيّر يعني إبقاء الارتباط الخطي المتعدد، ولا يوجد جواب تلقائي مباشرة عندما نبحث الارتباط الخطي المتعدد، وسنبحث هذا الموضوع فيما بعد.

5-3- طرق اكتشاف الارتباط الخطي المتعدد

كيف نعرف أن المعادلة تتضمن مشكلة الارتباط الخطي المتعدد؟ الخطوة الأولى للتعرف على وجود الارتباط الخطي المتعدد في المعادلة، علينا

أن نعلم أنه في الواقع العملي من غير الممكن أن نجد مجموعة متغيّرات تفسيرية غير مرتبطة ببعضها.

النقطة الثانية أن الارتباط الخطي المتعدد هو ظاهرة عينة sample، ومن الممكن أن يتغيّر من عينة لأخرى بالاعتماد على خصائص العينة، والأسس النظرية للمعادلة ليست مهمة في اكتشاف الارتباط الخطي المتعدد مثل اكتشاف حذف المتغيّرات أو عدم صحة شكل الدالة.

لأن الارتباط الخطي المتعدد ظاهرة عينة ومستوى الضرر لأثرها مشكلة قوة الارتباط، وسيتم استخدام عدة طرق لكشفه باختبارات ليس لها درجات حرجة أو مستوى معنوية، وسنفحص اثنين من أكثرها استخداماً لهذه الخصائص وهي:

5-3-1- معامل الارتباط البسيط

احد طرق اكتشاف قوة الارتباط الخطي المتعدد هو اختبار معاملات الارتباط الخطي البسيط بين المتغيّرات التفسيرية، فإذا كان r بالقيمة المطلقة مرتفعاً، يكون هذان المتغيّران مرتبطين تماماً ويكون الارتباط الخطي المتعدد مشكلة محتملة، والمشكلة هنا تحديد القيمة التي تعتبر كبيرة، ويعتبر أغلب الباحثين أن قيمة 0.9 بداية حدوث المشكلة، مثلاً كان في المعادلة (5.7) معامل الارتباط البسيط بين الدخل المتاح والأصول السائلة 0.986، وهذا المعامل مرتفعاً في معادلة فيها متغيّرين مستقلين. وبعض الباحثين يختار 0.80 ويأخذ بالاعتبار أن الارتباط الخطي المتعدد يحدث عندما تزيد القيمة المطلقة لمعامل الارتباط البسيط عن 0.80، والجواب الأفضل لاعتبار قيمة r مرتفعةً يكون بالنظر إلى آثاره، فإذا سبب زيادة غير مقبولة في تباين معاملات التقدير الذي نهتم فيه وأصبحت غير معنوية؛ هنا تظهر المشكلة.

لكن كن حذراً في استخدام معامل الارتباط البسيط كمؤشر على وجود الارتباط الخطي المتعدد في نموذج يتضمن أكثر من متغيّرين تفسيريين، وقد تسبب مجموعة متغيّرات مستقلة تعمل معاً الارتباط الخطي المتعدد دون أن يكون لأي معامل ارتباط بسيط درجة مرتفعة بما فيه الكفاية، وبالنتيجة يجب أن يكون معامل الارتباط البسيط اختباراً كافي وليس ضروري sufficient but not necessary للارتباط الخطي المتعدد، كذلك يشير ارتفاع r إلى احتمالية وجود ارتباط خطي متعدد وانخفاضه لا يعني إثبات غير ذلك.

5-3-2- عوامل تضخم التباين (VIFs)

أحد مقاييس خطورة الارتباط الخطي المتعدد سهل الاستخدام وله شهرة؛ إنه عامل تضخم التباين **Variance Inflation Factor (VIF)**؛ وهو أسلوب لاكتشاف خطورة الارتباط الخطي المتعدد بالنظر إلى مدى إمكانية تفسير أي متغيّر تفسيري بجميع المتغيّرات التفسيرية الأخرى في المعادلة، وذلك بحساب VIF لكل متغيّر تفسيري في المعادلة، وهو مقياس index يبين مقدار زيادة الارتباط الخطي المتعدد لتباين المعلمة المُقدّرة؛ حيث يشير ارتفاع VIF إلى أن الارتباط الخطي المتعدد يزيد من التباين المُقدّر للمعامل المُقدّر وينتج عنه انخفاض إحصائية t.

افرض أنك تريد استخدام VIF لاكتشاف الارتباط الخطي المتعدد في المعادلة الأصلية بمتغيّرات مستقلة عددها k:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + \varepsilon_i \qquad (5.9)$$

يتطلب إجراء حساب VIF عدده k مرة مختلفة؛ أي معامل واحد لكل X_i ويتكون حساب VIF لكل X_i من ثلاث خطوات كما يلي:

1- قدر انحدار المربعات الصغرى العادية يكون فيه المتغيّر X_i دالة في جميع المتغيّرات التفسيرية الأخرى في المعادلة، مثلاً إذا كان i=1 تكون هذه المعادلة:

$$X_1 = \alpha_1 + \alpha_2 X_2 + \alpha_3 X_3 + \cdots + \alpha_k X_k + \upsilon \qquad (5.10)$$

حيث أن υ حد الخطأ العشوائي، وأن X_1 غير موجود في الجانب الأيمن من المعادلة (5.10) التي يشار إليها بأنها انحدار مساعد auxiliary، وبالتالي يكون لدينا k انحداراً مساعداً لكل متغيّر مستقل في المعادلة الأصلية.

2- احسب عامل تضخم التباين للمعلمة $\hat{\beta}$:

$$VIF(\hat{\beta}_i) = \frac{1}{(1 - R_i^2)} \qquad (5.11)$$

حيث R_i^2 هو معامل تحديد (R^2 غير المصحح) المعادلة المساعدة في الخطوة الأولى.

3- حلل درجة الارتباط الخطي المتعدد بتقييم حجم $VIF(\hat{\beta}_i)$: فارتفاع قيمة VIF لمتغير يزيد من تباين معامل المتغيّر المُقدّر (مع بقاء تباين حد الخطأ ثابتاً)، وبالتالي يكون إرتفاع VIF أكبر تأثير للارتباط الخطي المتعدد.

نستطيع رؤية حالات درجة الارتباط المشترك intercorrelation بين المتغيّرات؛ فإذا ارتفع R_i^2 سترتفع VIF_i بمعدل متزايد، والاقتراب من اللانهاية في حالة الارتباط الخطي التام ($R_i^2 = 1$). وإذا كانت $R_i^2 = 0$ فإنها تشير إلى عدم وجود ارتباط خطي متعدد ويساوي ناتج VIF الواحد، ولا يوجد جدول قيم حرجة رسمي لقيم VIF، وكقاعدة عامة إذا كان $VIF(\hat{\beta}_i) > 5$ يكون الارتباط الخطي المتعدد قوياً، ويعرض الجدول أدناه قيم مختلفة لـ R_i^2 وما يقابلها من VIF_i.

R_i^2	VIF_i
0	1
0.5	2
0.8	5
0.9	10
0.95	20
0.975	40
0.99	100
0.995	200
0.999	1000

قيم VIF_i التي تتجاوز 10 تقدم برهاناً على وجود الارتباط الخطي المتعدد، ونرى من هذا الجدول أنه عندما تكون $R^2 > 0.9$، يوجد الارتباط الخطي غير التام؛ ويؤثر الارتباط الخطي على تباين مُقدّرات OLS، وعلى التباين المشترك، ومن هذه الحقيقة قد تظهر امكانية معلمات معكوسة الإشارة، ويمكن تلخيص نتيجة الارتباط الخطي المتعدد بما يلي:

1- قد يكون تقدير معاملات OLS غير دقيق؛ بمعنى أن ارتفاع الأخطاء المعيارية يؤدي إلى فترة ثقة أوسع.

2- المعاملات المتضررة قد تفشل من الحصول على دلالة إحصائية نتيجة انخفاض احصائية t التي تؤدي إلى اسقاط خاطئ لتأثير المتغيّر في نموذج الانحدار.

3- قد تكون اشارة المعاملات المُقدّرة معاكسة لما هو متوقع.

4- قد ينتج اضافة أو طرح مشاهدات قليلة تغيّراً كبيراً في المعاملات المُقدّرة.

مثال

يوجد لدينا ثلاث متغيّرات هي: Y و X_2 و X_3، وحيث يوجد ارتباط خطي بين X_2 و X_3 تم الحصول على مصفوفة ارتباط المتغيّرات الثلاثة وكانت كما يلي:

جدول (5-1) مصفوفة الارتباط

	Y	X2	X3
Y	1	0.8573	0.8574
X2	0.8573	1	0.9999
X3	0.8574	0.9999	1

النتائج متماثلة منتظمة، بينما عناصر القطر تساوي 1 لأنه معامل ارتباط لنفس المتغيّرات، ونستطيع رؤية أن Y له ارتباط موجب مرتفع مع X_2 و X_3، وكذلك X_2 و X_3 كأنهما نفس المتغيّرين (معامل الارتباط يساوي 0.999995)، ومن الواضح أنه يشتبه بوجود امكانية عالية لأثر سلبي للارتباط الخطي المتعدد.

قدرنا الانحدار بكلا المتغيّرين التفسيريين، وحصلنا على النتائج في الجدول (5-2)، ونرى أن أثر X_2 على Y سالب، وأثر X_3 موجب. وكلا المتغيّرين غير معنوي، وهذه النتيجة غريبة باعتبار أن لكلا المتغيّرين ارتباط كبير مع Y كما يظهر أعلاه، وبالتالي سيتم تقدير نموذج يتضمن X_2 فقط.

جدول (5-2) نتائج الانحدار(النموذج الكامل)

Dependent Variable: Y
Method: Least Squares
Date: 05/24/12 Time: 11:54
Sample: 1 25
Included observations: 25

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	35.86766	19.38717	1.850073	0.0778
X2	-6.326498	33.75096	**-0.187446**	**0.8530**
X3	1.789761	8.438325	0.212099	0.8340

R-squared	0.735622	Mean dependent var	169.3680
Adjusted R-squared	0.711587	S.D. dependent var	79.05857
S.E. of regression	42.45768	Akaike info criterion	10.44706
Sum squared resid	39658.40	Schwarz criterion	10.59332
Log likelihood	-127.5882	Hannan-Quinn criter.	10.48763
F-statistic	30.60702	Durbin-Watson stat	2.875574
Prob(F-statistic)	0.000000		

نعيد وصف المعادلة ونحذف المتغيّر X_3، ونحصل على النتائج المعروضة في جدول (5-3)، وسنرى أن X_2 موجب ومعنوي احصائياً (احصائية t تساوي 7.98).

جدول (5-3) نتائج الانحدار(حذف X_3)

Dependent Variable: Y
Method: Least Squares
Included observations: 25

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	36.71861	18.56953	1.977358	0.0601
X2	0.832012	0.104149	**7.988678**	**0.0000**

R-squared	0.735081	Mean dependent var	169.3680
Adjusted R-squared	0.723563	S.D. dependent var	79.05857
S.E. of regression	41.56686	Akaike info criterion	10.36910
Sum squared resid	39739.49	Schwarz criterion	10.46661
Log likelihood	-127.6138	Hannan-Quinn criter.	10.39615
F-statistic	63.81897	Durbin-Watson stat	2.921548
Prob(F-statistic)	0.000000		

ويتم اعادة تقدير النموذج ليتضمن فقط X_3 ونحصل على النتائج في الجدول (5-4)، ونرى أن X_3 معنوي جداً وموجب الاشارة.

جدول (5-4) نتائج الانحدار(حذف X_2)

Dependent Variable: Y
Method: Least Squares
Included observations: 25

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	36.60968	18.57637	1.970766	0.0609
X3	0.208034	0.026033	**7.991106**	**0.0000**

R-squared	0.735199	Mean dependent var	169.3680
Adjusted R-squared	0.723686	S.D. dependent var	79.05857
S.E. of regression	41.55758	Akaike info criterion	10.36866
Sum squared resid	39721.74	Schwarz criterion	10.46617
Log likelihood	-127.6082	Hannan-Quinn criter.	10.39570
F-statistic	63.85778	Durbin-Watson stat	2.916396
Prob(F-statistic)	0.000000		

أخيراً ننفذ انحداراً مساعداً للمتغيّر X_2 على المقطع c وعلى X_3 ونحصل على النتائج التي يبينها الجدول (5-5)، ونلاحظ أن قيمة احصائية t مرتفعة جداً (1521.542) بينما R^2 قريبة من 1.

جدول (5-5) نتائج الانحدار(انحدار X_2 على X_3)

Dependent Variable: X2
Method: Least Squares
Sample: 1 25
Included observations: 25

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.117288	0.117251	-1.000310	0.3276
X3	0.250016	0.000164	**1521.542**	**0.0000**

R-squared	**0.999990**	Mean dependent var	159.4320
Adjusted R-squared	0.999990	S.D. dependent var	81.46795
S.E. of regression	0.262305	Akaike info criterion	0.237999
Sum squared resid	1.582488	Schwarz criterion	0.335509
Log likelihood	-0.974992	Hannan-Quinn criter.	0.265045
F-statistic	2315090.	Durbin-Watson stat	2.082420
Prob(F-statistic)	0.000000		

ويمكن تلخيص نتيجة هذا التحليل كما يلي:

1- الارتباط بين المتغيّرات التفسيرية كان مرتفعاً جداً، ويبين وجود ارتباط خطي متعدد، وأشرنا في النظرية أن معاملات ارتباط المتغيّرات التفسيرية ليست كافية لكشف الارتباط الخطي المتعدد.

2- الخطأ المعياري أو نسبة احصائية t للمعاملات المُقدّرة تختلف من تقدير لتقدير، مبينة أن مشكلة الارتباط الخطي موجودة.

3- كان استقرار معاملات التقدير اشكالية كبيرة كذلك، ويتم تقدير معاملات مره سالبة ومرة موجبة لنفس المتغيّر في وصفين مختلفين.

4- يبين R^2 من الانحدار المساعد المرتفع وجود ارتباط خطي محتوم الأثر لتقديرنا.

5-4- علاج الارتباط الخطي المتعدد

ماذا عليك فعلة لتقليل نتائج الارتباط الخطي المتعدد القوي؟ لا يوجد جواب مباشر؛ لأن الارتباط الخطي المتعدد ظاهرة قد تتغيّر من عينة لأخرى لنفس وصف معادلة الانحدار، والهدف من هذا الجزء وضع عدة خيارات علاج ملائمة للارتباط الخطي المتعدد لنفس الظروف.

أ- **اسقاط المتغيّرات الزائدة:** الحل البسيط لتقليل نتائج الارتباط الخطي المتعدد هو اسقاط أحد متغيّرات انحدار الارتباط الخطي المتعدد؛ مثلاً بعض الباحثين يُدخل العديد من المتغيّرات في الانحدار، وبعض المتغيّرات يقيس نفس الأثر، وبعضها غير مؤثر، وبدلاً من هذه المتغيّرات المسماة بالمتغيّرات الزائدة قد يكفينا واحداً منها لبيان الأثر على المتغيّر التابع، مثلاً دالة الطلب الكلي ليس من المنطق إضافة الدخل المتاح والناتج المحلي الإجمالي؛ لأن كل منهما يقيس نفس الأثر وهو أثر الدخل، والخبير لا يدخل عدد السكان والدخل المتاح في نفس دالة الطلب الكلي لأن كل منهما يقيس أثر حجم السوق، واسقاط مثل هذه المتغيّرات الزائدة تعمل على تعويض خطأ الوصف، وهذا ما رأيناه في مثال الاستهلاك لطلاب فصل الاقتصاد القياسي كدالة في الدخل المتاح والأصول السائلة معاً، حيث أظهرت المعادلة أن معاملاتهما كانت غير معنوية، وعندما اسقطنا الأصول السائلة أصبح معامل الدخل المتاح معنوياً، وكذلك حينما قدرنا

الاستهلاك كدالة في الأصول السائلة كان معاملها معنوياً، فإسقاط أحد المتغيّرين أزال الارتباط الخطي المتعدد بين المتغيّرين التفسيريين، وفي هذه الحالة تدعم النظرية فرضية أن الدخل المتاح يحدد الاستهلاك أكثر من فرضية الأصول السائلة.

ب- تحويل المتغيّرات: في حالات نادرة تكون نتائج الارتباط الخطي المتعدد خطرة عندما تكون جميع المتغيّرات مهمّة اعتماداً على الخلفية النظرية، وفي هذه الحالة لا يساعدنا ابقائها كما هي، وبدلاً من اسقاط متغيّر من الممكن تحويل المتغيّرات في المعادلة للتخلص من الارتباط الخطي المتعدد على الأقل، وأكثر التحويلات شيوعاً هي:

1- مزج المتغيّرات.

2- تحويل متغيّرات المعادلة إلى الفرق الأول.

يتوافق أسلوب مزج متغيّرين أو أكثر بإنشاء متغيّر جديد تكون فيه دالة في المتغيّرات الأصلية وتحل المتغيّرات الجديدة محل المتغيّرات القديمة في معادلة الانحدار، لكن كن على حذر عند مزج المتغيّرات، وأن صنع المتغيّر الجديد له معنى بمفرده.

مثلاً إذا كان X_1 و X_2 لهما ارتباطاً خطياً مرتفعاً، يكون المتغيّر الجديد $X_3 = X_1 + X_2$ (أو أي مزيج خطي للمتغيّرين مثل $k_1X_1 + k_2X_2$) قد يحل محل المتغيّرين في إعادة تقدير النموذج، وهذا الأسلوب مفيد إذا أردنا استخدام المعادلة للتوقعات، ومن مساوئ هذا الأسلوب هو أن قسمي مزيج المتغيّر لها نفس المعامل في المعادلة المعاد تقديرها، مثلاً إذا كان $X_3 = X_1 + X_2$، فإن:

$$Y = \beta_1 + \beta_3 X_3 + \varepsilon = \beta_1 + \beta_3 (X_1 + X_2) + \varepsilon \quad (5.12)$$

والنوع الثاني: التحويل للفرق الأول، وهو تغيّر المتغيّر من الفترة الزمنية السابقة إلى الفترة الزمنية الحالية (الذي نشير إليه بدلتا delta أو Δ)، والذي نعرّفه كما يلي:

$$\Delta X_t = X_t - X_{t-1}$$

3- زيادة حجم العينة.

5-5- مثال كامل يبحث الارتباط الخطي المتعدد

سنتعامل مع مثال كامل يبحث تأثير الارتباط الخطي المتعدد لنموذج الطلب على السمك في الولايات المتحدة الأمريكية من 1946 إلى 1970، وافرض أنك قررت التأكد من فكرة قرار بابا الفاتيكان في عام 1966 بالسماح للكاثوليك بأكل اللحوم يوم الجمعة خلال فترة الصيام التي تسبق عيد الفصح أدى إلى إنتقال دالة الطلب على اللحوم، وكانت الدالة المفترضة كما يلي:

$$F_t = f(\overset{-}{PF_t}, \overset{+}{PB_t}, \overset{+}{Yd_t}, \overset{+}{N_t}, \overset{-}{P_t}) + \varepsilon_i \qquad (5.13)$$

حيث أن:

F_t : كمية السمك المستهلك لكل شخص في السنة t.

PF_t: الرقم القياسي لأسعار السمك في السنة t.

PB_t: الرقم القياسي لأسعار لحوم البقر في السنة t.

Yd_t: الدخل الحقيقي المتاح لكل شخص في السنة t.

N_t: عدد الكاثوليك في الولايات المتحدة (عشرات الآلاف) في السنة t.

P_t: متغيّر وهمي يساوي 1 لما بعد قرار البابا في عام 1966 وصفر لغيرها.

جدول (5-6) بيانات مثال الطلب على السمك

Year	F	PF	PB	N	Yd
1946	12.8	56.0	50.1	24402	1606
1947	12.3	64.3	71.3	25268	1513
1948	13.1	74.1	81.0	26076	1567
1949	12.9	74.5	76.2	26718	1547
1950	13.8	73.1	80.3	27766	1646
1951	13.2	83.4	91.0	28635	1657
1952	13.3	81.3	90.2	29408	1678
1953	13.6	78.2	84.2	30425	1726
1954	13.5	78.7	83.7	31648	1714
1955	12.9	77.1	77.1	32576	1795
1956	12.9	77.0	74.5	33574	1839
1957	12.8	78.0	82.8	34564	1844
1958	13.3	83.4	92.2	36024	1831
1959	13.7	84.9	88.8	39505	1881
1960	13.2	85.0	87.2	40871	1883
1961	13.7	86.9	88.3	42105	1909
1962	13.6	90.5	90.1	42882	1969
1963	13.7	90.3	88.7	43847	2015
1964	13.5	88.2	87.3	44874	2126
1965	13.9	90.8	93.9	45640	2239
1966	13.9	96.7	102.6	46246	2335
1967	13.6	100.0	100.0	46864	2403
1968	14.0	101.6	102.3	47468	2486
1969	14.2	107.2	111.4	47873	2534
1970	14.8	118.0	117.6	47872	2610

ويكون شكل المعادلة كما يلي:

$$F_t = \beta_0 + \beta_1 PF_t + \beta_2 PB_t + \beta_3 LnYd_t + \beta_4 N_t + \beta_5 P_t + \varepsilon \quad (5.14)$$

وبما انك توقعت إشارة سالبة للمعامل يجب أن تكون الفرضية الأساسية $H_0 : \beta_5 \geq 0$ ، وكذلك أختيار دالة شبه لوغاريتمية لربط الدخل المتاح بكمية السمك المستهلك، وهذا يتوافق مع النظرية. وعندما يزيد الدخل ستنخفض حصة الدخل الإضافي الموجه لاستهلاك السمك، وسنتحرى النموذج ونتائج الارتباط الخطي:

وبعد جمع البيانات (الجدول 5-6) سنحصل على تقدير OLS التالي:

Dependent Variable: F
Method: Least Squares
Date: 12/19/15 Time: 18:54
Sample: 1946 1970
Included observations: 25

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.988398	12.98474	-0.153134	0.8799
PF	0.039502	0.031015	1.273645	0.2182
PB	-0.000777	0.020200	-0.038453	0.9697
LOG(YD)	1.770237	1.872606	0.945333	0.3564
N	-3.14E-05	3.28E-05	-0.957715	0.3502
P	-0.355258	0.353120	-1.006054	0.3270

R-squared	0.735631	Mean dependent var	13.44800
Adjusted R-squared	0.666060	S.D. dependent var	0.533948
S.E. of regression	0.308555	Akaike info criterion	0.691730
Sum squared resid	1.808918	Schwarz criterion	0.984260
Log likelihood	-2.646625	Hannan-Quinn criter.	0.772865
F-statistic	10.57385	Durbin-Watson stat	2.214949
Prob(F-statistic)	0.000057		

هذه النتيجة غير مشجعة لأن المعاملات غير معنوية، وأن إشارة ثلاثة معاملات غير متوقعة، وقد تظهر هذه المشكلة إذا حذفنا بعض المتغيرات على سبيل المثال (تحيّز معلمات)، أو أن المتغيرات غير ملائمة، أو وجود ارتباط خطي متعدد.

من أين نبدأ؟ إذا كان لديك الثقة في استعراض الأدبيات الاقتصادية والأعمال النظرية قبل تقدير المعادلة، يكون من الأفضل رؤية فيما إذا كانت هناك إشارات للارتباط الخطي المتعدد، داعماً قرارك بقيمة $\overline{R}^2$ التي تساوي 0.666 الذي يبدو مرتفعاً بالنسبة لقيم t غير المعنوية.

أحد مقاييس الارتباط الخطي المتعدد هو حجم معاملات الارتباط البسيط، أنظر إلى المتغيّرات، فأي زوج أو مجموعة متغيّرات قد يكون ارتباطها معنوي؟ يظهر لنا الدخل المتاح للفرد وعدد الكاثوليك متماثلان ليكون الارتباط بينهما مرتفعاً، وكلاهما جزء من المعادلة لقياس قوة الشراء، وكان الارتباط بين N_t و Yd_t يساوي 0.946.

ولا يوجد سبب للتفكير بأن أسعار السلع قد تتحرك مع بعض، وبما أن الأسعار المشاهدة هي أسعار توازنية، فقد تؤثر صدمات العرض والطلب على الأرقام القياسية للحم البقر والأسماك بنفس الطريقة، مثلاً تسرب النفط يجعل الأسماك غير قابلة للتسويق وتجعل من المسلّم به رفع أسعار السمك، لكن هذا الارتفاع سيدفع الطلب على لحوم البقر للأعلى، وبالتالي زيادة سعر لحوم البقر، وتميل الأسعار البديلة للتحرك مع بعضها، وإذا عدنا لمعاملات الارتباط البسيط بين السعرين تكون 0.958، ومع

الارتباط الذاتي بين المتغيّرين بعكس الاشارة المتوقعة، ومع الارتباط الخطي المتعدد يكون توزيع $\hat{\beta}$ واسعاً ومن المحتمل أن نشاهد زيادة الإشارات غير المتوقعة.

الطريقة الثانية لاكتشاف الارتباط الخطي المتعدد هي حجم عوامل تضخم التباين، وكذلك يشير إلى خطورة المشكلة، ويتوقع من جميع VIF للمعادلة أن تكون $VIF > 5$ مشيراً لخطورة الارتباط الخطي المتعدد:

Variance Inflation Factors
Date: 12/19/15 Time: 18:57
Sample: 1946 1970
Included observations: 25

Variable	Coefficient Variance	Uncentered VIF	Centered VIF
C	168.6034	44273.21	NA
PF	0.000962	1857.920	**42.88122**
PB	0.000408	843.1285	**18.77374**
LOG(YD)	3.506653	52571.80	**23.50889**
N	1.08E-09	403.8103	**18.51974**
P	0.124694	5.238885	**4.400663**

تظهر النتائج وجود الارتباط الخطي المتعدد في النموذج، ماذا عليك عملة حول هذه النتائج؟ عليك العودة إلى المعادلة واستعرض تقديرها واستعرض المفاهيم النظرية.

تستطيع التغلب على الارتباط الخطي المتعدد بين الدخل وعدد الكاثوليك، وبالنتيجة عليك حذف المتغيّرات الزائدة، ولا ينبغي وجودها

مع بعض في نفس المعادلة، وبالنظر إلى أن المنطق وراء تضمين عدد الكاثوليك في معادلة الطلب على السمك ضعيف إلى حد ما، عليك أن تقرر اسقاط N ، وكانت النتيجة كما يلي:

Dependent Variable: F
Method: Least Squares
Date: 12/19/15 Time: 19:01
Sample: 1946 1970
Included observations: 25

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.961108	7.773354	1.024154	0.3180
PF	0.027993	0.028533	0.981075	0.3383
PB	0.004692	0.019336	0.242675	0.8107
LOG(YD)	0.360363	1.154974	0.312010	0.7583
P	-0.124462	0.257573	-0.483211	0.6342

R-squared	0.722869	Mean dependent var	13.44800
Adjusted R-squared	0.667443	S.D. dependent var	0.533948
S.E. of regression	0.307916	Akaike info criterion	0.658876
Sum squared resid	1.896243	Schwarz criterion	0.902651
Log likelihood	-3.235945	Hannan-Quinn criter.	0.726488
F-statistic	13.04200	Durbin-Watson stat	2.236966
Prob(F-statistic)	0.000022		

هل تظهر نتائج هذا التقدير أنه تم حل مشكلة الارتباط الخطي المتعدد؟ وباسقاط N تم إزالة متغير زائد من المعادلة، إلا أن خطورة الارتباط الخطي المتعدد مقاساً بأساليب الكشف، مع بقاء ظهور الارتباط الخطي المتعدد الذي يشتمل على متغيري السعر، ماذا علينا فعله؟

في حالة الأسعار لا نملك خياراً لاسقاط أي منها؛ لأن PB و PF مهمّين نظرياً في النموذج. وهذه الحالة تستحق تحقيق علاج آخر محتمل، وهو تحويل المتغيّرات، فكانت إحدى البدائل لتكوين تحويل متغيّري السعرين بقسمة أحدهما على الاخر لتشكيل متغيّر سعر نسبي:

$$RP_t = \frac{PF_t}{PB_t}$$

افرض أننا قررنا تقدير المعادلة الأخيرة:

$$F_t = f(\overset{-}{RP_t}, \overset{+}{Yd_t}, \overset{-}{P_t}) + \varepsilon$$

وحصلنا على:

Dependent Variable: F
Method: Least Squares
Date: 12/19/15 Time: 19:04
Sample: 1946 1970
Included observations: 25

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-5.168676	4.832730	-1.069515	0.2970
PF/PB	-1.930897	1.430728	-1.349591	0.1915
LOG(YD)	2.711743	0.656781	4.128838	0.0005
P	0.005197	0.280080	0.018554	0.9854

R-squared	0.639721	Mean dependent var	13.44800
Adjusted R-squared	0.588252	S.D. dependent var	0.533948
S.E. of regression	0.342621	Akaike info criterion	0.841263
Sum squared resid	2.465174	Schwarz criterion	1.036284
Log likelihood	-6.515793	Hannan-Quinn criter.	0.895354
F-statistic	12.42938	Durbin-Watson stat	1.597750
Prob(F-statistic)	0.000069		

أصبحت المعادلة لا تعاني من مشكلة الارتباط الخطي المتعدد، وهذا ما تظهره قيم *VIF* وجميعها أقل من 3، وهذا ما بينته النتائج التالية:

Variance Inflation Factors
Date: 12/19/15 Time: 21:06
Sample: 1946 1970
Included observations: 25

Variable	Coefficient Variance	Uncentered VIF	Centered VIF
C	23.35528	4973.898	NA
PF/PB	2.046983	409.7554	**1.083755**
LOG(YD)	0.431361	5244.913	**2.345404**
P	0.078445	2.672978	**2.245302**

وإذا أردنا اختبار الفرضية الأساسية لعدم الأثر فقد تبين عدم قدرتنا على رفضها، وتظهر أن قرار بابا الفاتيكان لم يخفض استهلاك السمك (المعامل غير معنوي).

تمارين

5-1- عرّف الارتباط الخطي المتعدد Multicollinearity، واشرح كيفية الكشف عن وجوده في الانحدار المُقدّر البسيط.

5-2- تعلم أن $\hat{\beta}=(X'X)^{-1}X'Y$، ماذا يحدث للمعلمة $\hat{\beta}$ عند حدوث ارتباط خطي تام بين المتغيّرات المستقلة X_i؟ وكيف تعرف وجود ارتباط خطي تام؟

5-3- اشرح ما هو VIF، وماذا يستخدم؟

5-4- يعتقد الباحث المبتدء بوجود ارتباط خطي متعدد عندما يتم تضمين المعادلة من غير قصد متغيّرين تفسيريين أو أكثر يخدمان نفس الهدف أو تقيس نفس الشيء، ما هي أزواج المتغيّرات التي في المعادلة تعتبر متغيرات زائدة؟

أ) GDP و NDP في معادلة اقتصاد كلي.

ب) سعر الثلاجة وسعر الغسالة في دالة الطلب على السلع المعمّر.

ج) عدد الدونمات المحصودة ومقدار السماد المستخدم، في دالة عرض المزروعات.

د) سعر الفائدة طويل الأجل وعرض النقود في دالة الاستثمار.

5-5- حاول أحد الباحثين تقدير دالة الطلب على الأصول التي تتضمن

ثلاثة متغيّرات تفسيرية هي: الثروة الحالية W_t، والثروة في الربع السابق W_{t-1}، والتغيّر في الثروة $\Delta W_t = W_t - W_{t-1}$. ما هي المشكلة التي تواجه الباحث؟ ماذا عليه فعله لحل هذه المشكلة؟

5-6- قُدّر انحدار لبيانات مقطعية تخص 44 ولاية لفهم نفقات الدفاع الاتحادية (الانحراف المعياري بين القوسين)

$$\hat{S}_i = -148.0 + \underset{(0.027)}{0.841}\, C_i - \underset{(0.1664)}{0.0115}\, P_i - \underset{(0.0092)}{0.0078}\, E_i$$

حيث أن:

S_i: النفقات السنوية (مليون دولار) على الدفاع في الولاية i.

C_i: العقود الممنوحة للخدمة العسكرية (المجندين) في السنة.

P_i: الفاتورة السنوية (مليون دولار) للعمال في الصناعات الموجهة للدفاع في الولاية i.

E_i: عدد الأفراد المدنيين العاملين في الصناعات الموجهة للدفاع في الولاية i.

أ) احسب إحصائية t واختبر الفرضيات عند مستوى معنوية 5٪، علماً بأن القيمة الجدولية عند مستوى معنوية 0.05 ودرجات حرية 40 تساوي 2.021.

ب) كان VIF لهذه المعادلة أكبر من 20، ولكل من C_i و P_i كان أكبر من 30، ماذا تستنتج من هذه المعلومات.

ج) ما هي مقترحاتك لاعادة تقدير هذه المعادلة بوصف مختلف؟ اشرح اجابتك.

5-7- افرض أن صديقك أعد نموذجاً لأثر الدخل على الاستهلاك في نموذج فصلي (ربعي) واكتشف أن الدخل يؤثر على الاستهلاك الدائم سنة على الأقل، وبالنتيجة قدّر صديقك النموذج التالي:

$$C_t = \beta_1 + \beta_2 Yd_t + \beta_3 Yd_{t-1} + \beta_4 Yd_{t-2} + \beta_5 Yd_{t-3} + u_i$$

أ) هل تتضمن هذه المعادلة ارتباطاً خطياً متعدداً تاماً؟

ب) هل تتضمن هذه المعادلة ارتباطاً خطياً غير متعدد تام؟

ج) ماذا علينا عمله لإزالة الارتباط الخطي المتعدد من هذه المعادلة؟ (إحدى الإجابات تقدير معادلة الانحدار الذاتي للفجوات الزمنية الموزعة ARDL في الفصل العاشر).

الفصل السادس

اختلاف التباين

Heteroskedasticity

هذه المشكلة تنتهك الفرضية (5) التي تنص على أن مشاهدات حد الخطأ لها تباين ثابت، وهذه الفرضية غير واقعية لمشاهدات حد الخطأ المتجانسة، وبشكل عام فإن اختلاف التباين ينتشر في البيانات المقطعية أكثر من بيانات السلاسل الزمنية؛ وهذا لا يعني أن السلاسل الزمنية تخلوا من هذه المشكلة، وهذا يتضح في دراسات السلاسل الزمنية للأسواق المالية.

وفي هذا الفصل سنجيب على الأسئلة الأربعة لعدم ثبات التباين (عدم التجانس) التي أشرنا إليها في الفصل السابق حول الارتباط الخطي المتعدد التالية:

1- ما هي طبيعة المشكلة؟

2- ما هي نتائج المشكلة؟

3- كيف نشخص المشكلة؟

4- ما هي علاجات المشكلة المتاحة؟

6-1- طبيعة مشكلة اختلاف التباين

سنبدأ بتعريف كلمة Homoskedasticity و Heteroskedasticity، فبعض المؤلفين يكتبها Homoscedasticity وبعضهم يستخدم Homoskedasticity بالاعتماد على أصلها اللاتيني، ويتكون كل من الكلمتين من جزئين: الجزء الأول من الكلمة اللاتينية **Homo** (التي تعني نفسه أو يساوي) أو **Hetero** (التي تعني مختلف أو لا يساوي)، والجزء الثاني من الكلمة اللاتينية **skedastic** (التي تعني انتشار أو تبعثر)؛ وعليه تعني Homoscedasticity انتشار متساوي، وبالجانب الآخر تعني Heteroskedasticity انتشار غير متساوٍ. ونستخدم التباين في القياس الاقتصادي كمقياس للانتشار، وبالتالي تبحث Heteroskedasticity عدم تساوي التباين أو اختلاف التباين.

وتذكيراً بفرضيات نموذج الانحدار الخطي التقليدي في الفصل الثاني والثالث، يجب أن يكون تباين حد الخطأ ثابتاً (متساوياً)، ونعرضه رياضياً كما يلي:

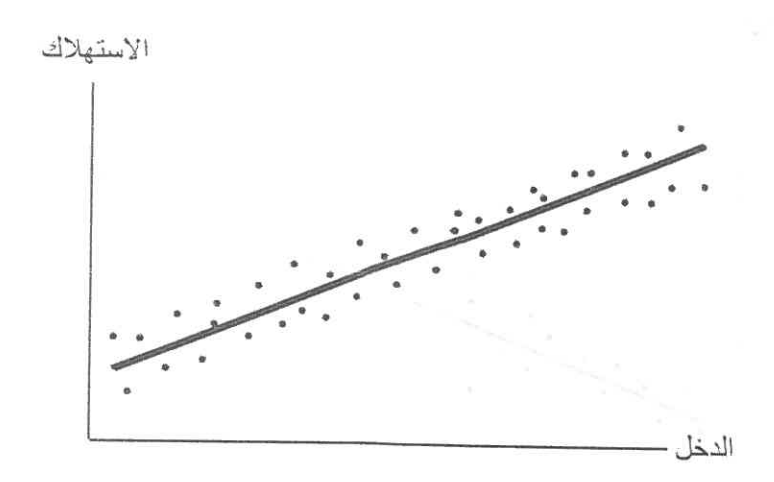
$$\operatorname{var}(u_i) = \sigma^2 \qquad (6.1)$$

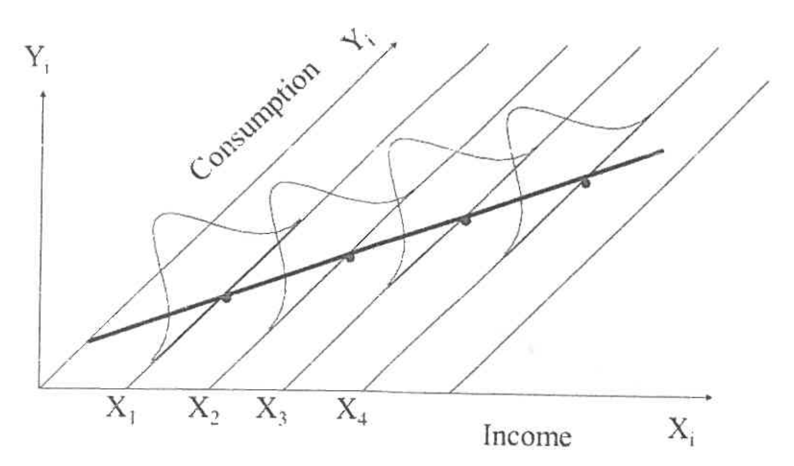
وبالتالي يعني وجود تباينٍ متساوي أن حدود الخطأ متساوية الانتشار Homoscedasticity. وبشكل عام يحدث عدم تساوي التباين في إطار البيانات المقطعية؛ ولا يعني هذا أنه غير ممكن في نماذج السلاسل الزمنية. وفي هذه الحالة تنتهك فرضية ثبات التباين، ويعتمد تباين حدود الخطأ على المشاهدات كما في المثال التالي:

$$\operatorname{var}(u_i) = \sigma_i^2 \qquad (6.2)$$

مع ملاحظة أن الفرق بين المعادلتين (6.1) و (6.2) هو الحرف المنخفض i المتعلق بالتباين σ^2، الذي يعني أن التباين يختلف باختلاف المشاهدة في العينة $i = 1, 2, \ldots, n$. ولمزيد من التوضيح نعود إلى شكل نموذج الانحدار البسيط بمتغيّرين:

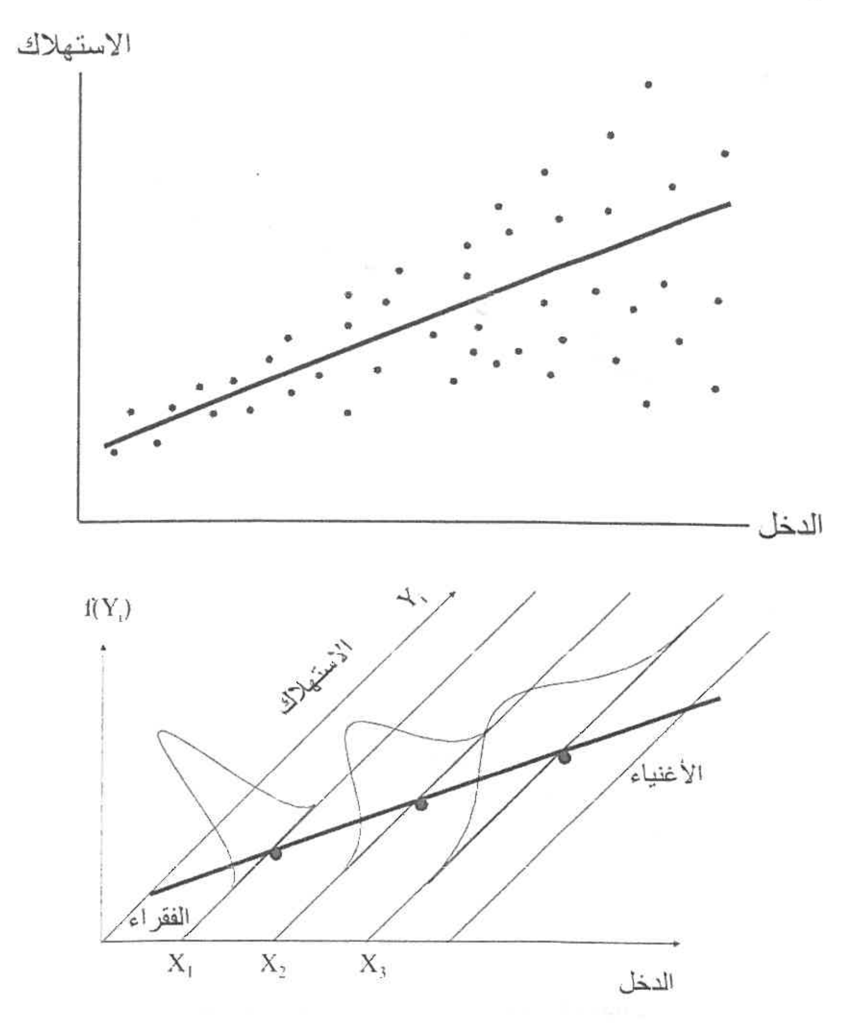
$$Y_i = \beta_1 + \beta_2 X_i + u_i \qquad (6.3)$$





شكل رقم 6-1: ثبات تباين البيانات

انظر صورة الانتشار لخط الانحدار للمجتمع في الشكل (6-1) وقارن بالشكل (6-2)، حيث تشير النقاط X_1 و X_2 و X_3 في الشكل (6-1) إلى قيم مختلفة من X، وأن $X_1 < X_2 < X_3$ التي تؤثر على Y تبيّن أنها قريبة من خط الانحدار وبانتشارٍ متساوٍ فوق وتحت خط الانحدار (أي أن الانتشار متساوي=Homoscedasticity).

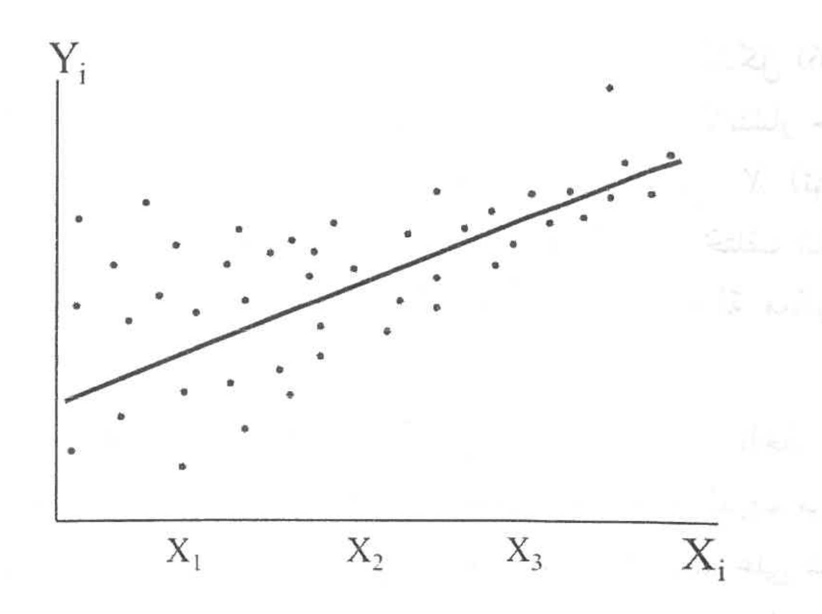


شكل رقم 6-2: مثال Heteroskedasticity بتباين متزايد

من جهة أخرى، تشير النقاط X_1 و X_2 و X_3 في الشكل (6-2) إلى اختلاف قيم X، ومن الواضح أن زيادة قيم X تزيد الانتشار حول الخط، وفي هذه الحالة، فإن الانتشار مختلف وغير متساوٍ لكل X_i (تؤخذ من الخط فوق وأسفل خط الانحدار)، وبالتالي لدينا تباين مختلف القيمة Heteroskedasticity، ومن الواضح أن الشكل (6-2) حالة معكوسة (انخفاض X تعني تباين أكبر).

كمثال على حالة اختلاف التباين الشكل (6-2)، نأخذ نمط الاستهلاك والدخل؛ فالافراد منخفضي مستوى الدخل ليس لديهم مرونة في انفاق دخلهم، أما الأفراد مرتفعي الدخل سينفقوا دخلهم على شراء الطعام والملابس والمواصلات و ... . مقارنة بمنخفضي الدخل، فإن نمط استهلاكهم لا يختلف كثيراً وسيكون الانتشار كبيراً أو منخفضاً قليلاً. ومن جانب آخر، لدى الافراد الاغنياء خيار واسع ومرونة في انفاقهم، وبعضهم يستهلك كثيراً وبعضهم يدخر كثيراً أو يستثمر في السوق المالي، وهذا يعني أن معدل الاستهلاك (حسب خط الانحدار) قد يختلف عن الاستهلاك الفعلي؛ لذا، سيكون انتشار مرتفعي الدخل مرتفعاً بأكثر من منخفضي الدخل.

ومثال على الحالة المتناقصة التي يصورها الشكل (6-3) هي البنوك الكبيرة التي لديها عمليات تجهيز بيانات متطورة ولها القدرة على الحساب بأقل وقت مقارنة مع البنوك الأصغر التي ليس لديها هذه الامكانيات، أو نماذج التعلم من الخطأ حيث تنخفض الخبرة تكون فرصة الاخطاء كبيرة (متغيّر درجة الاداء Y للاختبار ومتغيّر الوقت X الذي أخذه الفرد في الاختبار سابقاً، أو ساعات إعداد الاختبار؛ فزيادة X تقلل تباين Y).



شكل رقم 6-3: مثال Heteroskedasticity بتباين متناقص

6-2- نتائج اختلاف التباين

نأخذ نموذج الانحدار الخطي التقليدي:

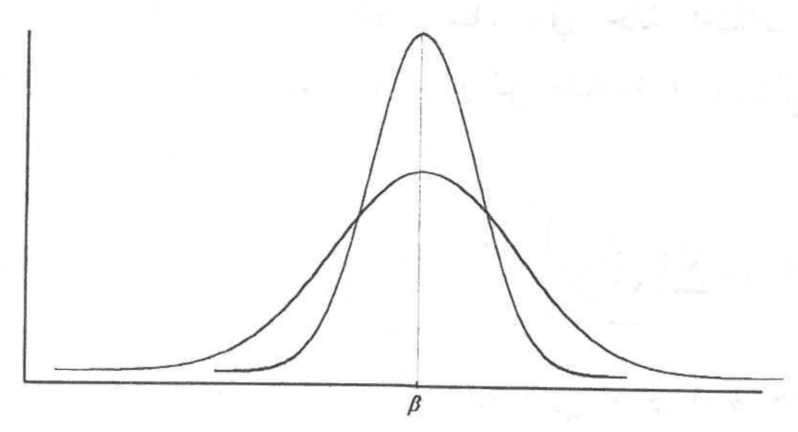
$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_k X_{ki} + u_i \qquad (6.4)$$

إذا كان حد الخطأ u_i معروفاً وتباينه غير ثابت في هذه المعادلة، فإننا نستطيع تلخيص نتائج مقدرات المربعات الصغرى $\hat{\beta}_i$ كما يلي:

1- تبقى معلمات OLS غير منحازة unbiased ومتسقة consistent؛ لأنه لا يوجد ارتباط بين أي متغيّر تفسيري وحد الخطأ، والمعادلة محددة بشكل صحيح إلا أنها تعاني من وجود مشكلة اختلاف التباين، وبالتالي ستعطي قيم $\hat{\beta}_i$ جيدة نسبياً.

2- يؤثر اختلاف التباين على توزيع $\hat{\beta}_i$ ويزيد تباين التوزيع ويجعل مقدرات طريقة OLS غير فعّالة؛ لأنه ينتهك خاصية تقليل التباين. ولفهم هذا انظر الشكل (6-4) الذي يبين توزيع معلمات $\hat{\beta}$ بوجود أو بعدم وجود اختلاف في التباين، ومن الواضح أن اختلاف التباين لا يسبب التحيّز لأن $\hat{\beta}$ تتركز حول $\hat{\beta}$ $\left(E(\hat{\beta})=\hat{\beta}\right)$، إلا أن اتساع التوزيع يجعله غير فعّال كثيراً، وبالتالي تكون مقدرات OLS ليست فعّالة كثيراً.

3- يؤثر اختلاف التباين كذلك على تباين $\hat{\beta}$ المقدرة (وبالتالي على الخطأ المعياري)، وفي الحقيقة يسبب وجود اختلاف التباين على منهجية OLS بتباين أقل من التقدير underestimate (والخطأ المعياري)، وبالتالي يؤدي إلى الحصول على قيم t أكبر من المتوقع وكذلك قيم F، وبالتالي فإن اختلاف التباين له تأثير واسع على اختبار الفرضيات، وإحصائية t و F تكون موثوقة لاختبار أي فرضية لأنها تؤدي إلى رفض الفرضية الأساسية.



شكل رقم 6-4: أثر اختلاف التباين على المعلمات المقدرة

ولنرى كيفية تأثير اختلاف التباين على مقدرات OLS، سنرى في البداية ماذا سيحدث لنموذج الانحدار البسيط، وبالتالي بيان أثر اختلاف التباين من شكل مصفوفة التباين- التباين المشترك لحدود الخطأ لنموذج الانحدار المتعدد، وبعد ذلك نرى الأثر باستخدام جبر المصفوفات في إطار الانحدار المتعدد.

سيتأثر تباين معامل الميل باختلاف التباين في نموذج الانحدار الخطي البسيط بمتغيّر تفسيري واحد وثابت، مذكراً بمعادلة تباين معامل $\hat{\beta}$ التالية:

$$\operatorname{var}(\hat{\beta}) = \sum\left(\frac{x_i}{\sum x_i^2}\right)^2 \sigma^2$$

$$= \frac{\sum x_i^2 \sigma^2}{\left(\sum x_i^2\right)^2} = \sigma^2 \frac{1}{\sum x_i^2} \qquad (6.5)$$

حيث أن $x_i = X_i - \bar{X}$، هذا في حالة تجانس تباين (homoskedasticity) حد الخطأ فقط، وفي حالة اختلاف التباين (heteroskedasticity) يتغيّر التباين مع كل مشاهدة i، وبالتالي يكون تباين $\hat{\beta}$ كما يلي:

$$\operatorname{var}(\hat{\beta}) = \sum\left(\frac{x_i}{\sum x_i^2}\right)^2 \sigma^2 = \frac{\sum x_i^2 \sigma_i^2}{\left(\sum x_i^2\right)^2} \qquad (6.6)$$

وهي تختلف عن (6.5)، والآن علينا شرح التحيّز الذي يحدث من وجود اختلاف التباين، فإذا وجد اختلاف التباين نحسب تباين $\hat{\beta}$ حسب

صيغة OLS المعيارية (6.5) بدلاً من المعادلة الصحيحة (6.6)، وسيكون التباين حتماً أقل من قيمة التباين الصحيحة والخطأ المعياري للمعلمة $\hat{\beta}$، وبالتالي سيكون لدينا نسبة t خاطئة كثيراً، وعدم الصحة تؤدي إلى استنتاج أن المتغيّر التفسيري X سيكون معنوياً احصائياً، بينما في الحقيقة يكون أثره على Y صفراً، وكذلك فترة الثقة حول β تكون أضيق من القيمة الصحيحة، ويؤدي إلى الاعتقاد بوجود دقة عالية في تقديرنا أكبر من الواقع وحالة مبررة احصائياً.

ومن المفيد أن نرى كيفية تأثير اختلاف التباين على شكل مصفوفة التباين-التباين المشترك لحدود الخطأ في نموذج الانحدار الخطي المتعدد التقليدي.

لنتذكر أن مصفوفة التباين-التباين المشترك للأخطاء هي:

$$E(uu')=\begin{pmatrix} \sigma^2 & 0 & 0 & \cdots & 0 \\ 0 & \sigma^2 & 0 & \cdots & 0 \\ 0 & 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma^2 \end{pmatrix}=\sigma^2 I_n$$

حيث أن I_n المصفوفة المحايدة $n\times n$.

أما في حالة وجود اختلاف التباين تصبح مصفوفة التباين-التباين المشترك للبواقي كما يلي:

$$E(uu')=\begin{pmatrix} \sigma_1^2 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & 0 & \cdots & 0 \\ 0 & 0 & \sigma_3^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma_n^2 \end{pmatrix}=\Omega \qquad (6.7)$$

تذكر أن مصفوفة التباين-التباين المشترك لمقدرات المربعات الصغرى $\hat{\beta}$ تعطي بـ:

$$cov(\hat{\beta})=E\left[(\hat{\beta}-\beta)(\hat{\beta}-\beta)'\right]$$

$$=E\left\{\left[(X'X)^{-1}X'u\right]\left[(X'X)^{-1}X'u\right]'\right\}$$

وحيث أن $(AB)' = \beta' A'$ فإن:

$$=E\left\{(X'X)^{-1}X'uu'X(X'X)^{-1}\right\}$$

$$=(X'X)^{-1}X'E(uu')X(X'X)^{-1}$$

$$=(X'X)^{-1}X'\Omega X(X'X)^{-1} \qquad (6.8)$$

وهي تختلف عن الصيغة التقليدية $\sigma^2(X'X)^{-1}$.

6-3- طرق الكشف عن اختلاف التباين

هناك طريقتان لبيان وجود اختلاف التباين: الأولى بمعاينة عدم تماثل الأشكال التي تسمى الطريقة غير الرسمية، بينما الثانية تطبيق الاختبار المناسب للكشف عن اختلاف التباين الذي يتضمن عدة اختبارات عن وجود اختلاف التباين.

أ) اختبار Breusch- Pagan LM test

طور Breusch and Pagan (1979) اختبار مضاعف لاغرانج (LM test) للكشف عن اختلاف التباين في النموذج التالي:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + ... + \beta_k X_{ki} + u_i \qquad (6.9)$$

حيث $\text{var}(u_i) = \sigma_i^2$، ويتضمن اختبار Breusch- Pagan الخطوات التالية:

1- نقدر نموذج (6.9) ونحصل على بواقي $(\hat{u}_i)$ معادلة هذا الانحدار.

2- نقدر الانحدار المساعد التالي:

$$\hat{u}_i^2 = a_1 + a_2 Z_{2i} + a_3 Z_{3i} + ... + a_p Z_{pi} + v_i \qquad (6.10)$$

حيث Z_{pi} هو مجموعة المتغيّرات التي نعتقد أنها تحدد تباين حد الخطأ (عادة يستخدم نيابة عن Z_{pi} المتغيّرات التفسيرية

المستخدمة في معادلة الانحدار الأصلية؛ أي X_S). (وتستخدم $\hat{u}_i^2$ بديلاً عن σ^2).

3- يتم صياغة الفرضية الأساسية والفرضية البديلة، وتكون الفرضية الأساسية لاختلاف التباين كما يلي:

$$H_0 : a_1 = a_2 = \cdots = a_p = 0 \qquad (6.11)$$

بينما تكون الفرضية البديلة وجود أحد a على الأقل يختلف عن الصفر.

4- احسب احصائية $LM = NR^2$، حيث N عدد المشاهدات المستخدمة في تقدير الانحدار المساعد في الخطوة (2)، و R^2 معامل التحديد لهذا الانحدار، وتتبع احصائية LM توزيع χ^2 بدرجات حرية $p-1$.

5- ترفض الفرضية الأساسية ويستنتج وجود دليل معنوي على اختلاف التباين عندما تكون احصائية LM اكبر من القيمة الحرجة $\left(\text{LM}-\text{احصائية} > \chi^2_{p-1,\alpha}\right)$، أو نحسب قيمة p-value ونرفض الفرضية الأساسية إذا كانت p-value أقل من مستوى معنوية α وعادة ما تكون $\alpha = 0.05$).

لحساب احصائية LM نحتاج حساب $LM = N * R^2$، حيث N عدد المشاهدات، و R^2 معامل التحديد للانحدار المساعد. وبعد ذلك نقارن احصائية LM بقيمة LM الحرجة ويتم الاستنتاج.

مثال

إذا أردنا اختبار اختلاف التباين Heteroskedasticity، حسب طريقة Breusch- Pagan ستتبع الخطوات التالية:

1- نقدر انحدار القيمة المضافة في الصناعة $(MANU)$ على الناتج المحلي الاجمالي (GDP) باستخدام بيانات الجدول (6-1) للنموذج التالي:

$$MANU = \beta_1 + \beta_2 GDP + u_i$$

$$MANU = -836.2304 + 0.0998\ GDP$$

ثم نحصل على بواقي $(\hat{u}_i)$ معادلة هذا الانحدار.

2- نقدر الانحدار المساعد التالي:

$$\hat{u}_i^2 = 9336 + 102.9405\ GDP$$

Dependent Variable: **U^2** $(\hat{u}_i^2)$
Method: Least Squares
Date: 12/23/15 Time: 11:33
Sample: 1 20
Included observations: **20**

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	9336466.	12718691	0.734074	0.4724
GDP	102.9405	84.07357	1.224410	0.2366

R-squared	**0.076884**	Mean dependent var	19869636
Adjusted R-squared	0.025600	S.D. dependent var	42441762
S.E. of regression	41894985	Akaike info criterion	38.03387
Sum squared resid	3.16E+16	Schwarz criterion	38.13344
Log likelihood	-378.3387	Hannan-Quinn criter.	38.05331
F-statistic	1.499179	Durbin-Watson stat	2.022928
Prob(F-statistic)	0.236582		

3- الفرضية الأساسية والفرضية البديلة لاختلاف التباين كما يلي:

$$H_0 : a_1 = a_2 = 0$$

بينما تكون الفرضية البديلة وجود أحد a على الأقل يختلف عن الصفر.

4- نحسب احصائية $LM = N R^2$

$$LM = 20 \times (0.076884) = 1.53768$$

5- وبما أن $\chi^2{}_{(1,0.05)} = 3.84$ ، فإن $LM = 1.53768 \; < \; \chi^2{}_{(1,0.05)} = 3.84$ وبالتالي نقبل الفرضية الأساسية ونستنتج عدم وجود دليل معنوي على اختلاف التباين عند مستوى معنوية $\alpha = 0.05$ ؛ أي أن تباين الخطأ العشوائي ثابت خلال العينة.

ب) اختبار Glesjer LM test

اشتمل اختبار Glesjer (1969) على الخطوات التالية وهي نفس خطوات اختبار Breusch- Pagan باستثناء الخطوة الثانية حيث يتكون الانحدار المساعد من معادلة أخرى:

1- نقدر نموذج (6.9) ونحصل على بواقي $(\hat{u}_i)$ معادلة هذا الانحدار.

2- نفذ الانحدار المساعد التالي:

$$|\hat{u}_i| = a_1 + a_2 Z_{2i} + a_3 Z_{3i} + \ldots + a_p Z_{pi} + v_i \qquad (6.12)$$

جدول (6-1) القيمة المضافة للصناعة التحويلية والناتج المحلي الاجمالي والسكان لعينة الدول العربية (القيمة بالمليون دولار)، 2010

الدولة	القيمة المضافة للصناعة MANU	الناتج المحلي الاجمالي GDP	بواقي الانحدار المساعد $\hat{u}_i$	$\hat{u}_i^2$
الأردن	4437	26970	-5459556	29806751717136
الامارات	28935	303812	-40281021	1622560652802440
البحرين	3923	20725	-4238365	17963737873225
تونس	6602	42109	-3223733	10392454455289
الجزائر	8036	161736	27048657	731629845503649
جيبوتي	30	1184	-8898890	79190243232100
السعودية	44757	457058	-56382747	3179014159266010
السودان	5904	69568	-16454384	270746752819456
سورية	2591	55621	-10532167	110926541715889
العراق	3300	123147	44638858	1992627643544160
عُمان	6170	60338	-14588303	212818584419809
قطر	9403	122903	-17839997	318265492960009
القمر	24	560	-8747220	76513857728400
الكويت	6623	132065	9922615	98458288438225
لبنان	3007	39873	-13421577	180138729166929
ليبيا	3451	74763	-6918100	47860107610000
مصر	35166	225339	150000000	22500000000000000
المغرب	12909	96805	-2687478	7222538000484
موريتانيا	123	3701	-9369864	87794351378496
اليمن	2291	28181	-12139823	147375302471329

3- يتم صياغة الفرضية الأساسية والفرضية البديلة، وتكون الفرضية الأساسية لاختلاف التباين كما يلي:

$$H_0 : a_1 = a_2 = \cdots = a_p = 0 \qquad (6.13)$$

بينما تكون الفرضية البديلة وجود أحد a على الأقل يختلف عن الصفر.

4- احسب احصائية $LM = NR^2$، حيث N عدد المشاهدات المستخدمة في تقدير الانحدار المساعد في الخطوة (2)، و R^2 معامل التحديد لهذا الانحدار، وتتبع احصائية LM توزيع χ^2 بدرجات حرية $p-1$.

5- ترفض الفرضية الأساسية ويستنتج وجود دليل معنوي لاختلاف التباين عندما تكون احصائية LM أكبر من القيمة الحرجة ($\text{LM}-\text{احصائية} > \chi^2_{p-1,\alpha}$)، أو حساب قيمة p-value وترفض الفرضية الأساسية إذا كانت قيمة p-value أقل من مستوى معنوية α التي هي عادة ما تكون ($\alpha = 0.05$).

لحساب احصائية LM نحسب $LM = N * R^2$، حيث N عدد المشاهدات، و R^2 معامل التحديد للانحدار المساعد. وبعد ذلك نقارن LM الحرجة باحصائية LM والاستنتاج.

مثال

نعيد تطبيق بيانات المثال السابق لإجراء اختبار Glesjer (1969)

1- نقدر نموذج (6.9) ونحصل على بواقي ($\hat{u}_i$) معادلة هذا الانحدار كما في الجدول (6-1).

2- نفذ الانحدار المساعد التالي:

$$|\hat{u}_i| = 2333.06335181 + 0.00603442855937\ GDP$$

3- الفرضية الأساسية والفرضية البديلة لاختلاف التباين:
$H_0 : a_1 = a_2 = 0$، بينما تكون الفرضية البديلة وجود أحد a على الأقل يختلف عن الصفر.

4- نحسب احصائية $LM = N R^2$ وكانت كما يلي:

$$LM = 20 \times 0.040497 = 0.80994$$

5- وبما أن $\left(LM = 0.80994 < \chi^2_{(1,0.05)} = 3.84\right)$ نقبل الفرضية الأساسية؛ أي أن تباين البواقي متجانس عند مستوى معنوية 0.05، وهذا يؤكد نتيجة الاختبار السابق.

ج) اختبار Harvey-Godfrey LM test

طور Harvey (1976) و Godfrey (1978) اختبار اختلاف التباين حسب الخطوات السابقة باستثناء الخطوة الثانية وكانت على النحو التالي:

$$\ln\left(\hat{u}_i^2\right) = a_1 + a_2 Z_{2i} + a_3 Z_{3i} + \ldots + a_p Z_{pi} + v_i \qquad (6.14)$$

ونقارن احصائية LM التي تحسب $LM = N * R^2$ للانحدار المساعد بالقيمة الحرجة لاحصائية LM كما سبق أعلاه.

فهل هذا الاختبار يؤكد النتائج السابقة؟

قدرنا المعادلة (6.14) وكانت النتائج كما يلي:

Dependent Variable: **LOG(U^2)**
Method: Least Squares
Date: 12/23/15 Time: 12:44
Sample: 1 20
Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	14.86978	0.895480	16.60537	0.0000
GDP	-4.02E-06	5.92E-06	-0.678826	0.5059

R-squared	**0.024961**	Mean dependent var	14.45862
Adjusted R-squared	-0.029208	S.D. dependent var	2.907529
S.E. of regression	2.949685	Akaike info criterion	5.095913
Sum squared resid	156.6115	Schwarz criterion	5.195486
Log likelihood	-48.95913	Hannan-Quinn criter.	5.115351
F-statistic	0.460805	Durbin-Watson stat	1.728171
Prob(F-statistic)	0.505883		

ثم قدرنا احصائية $LM = N * R^2$ للانحدار المساعد، وكانت النتيجة كما يلي:

$$LM = 20 \times 0.024961 = 0.49922$$

ونقارن هذه القيمة بالقيمة الحرجة لكاي تربيع، وبما أن قيمة 0.49922 أصغر من قيمة χ^2 الحرجة، سنقبل الفرضية الأساسية القائلة بتجانس تباين حد الخطأ.

د) اختبار Park LM test

طور Park (1966) اختباراً بديلاً لاختبار LM تضمن نفس الخطوات في الاختبارات السابقة باستثناء الخطوة الثانية التي كانت كما يلي:

$$\ln\left(\hat{u}_i^2\right) = a_1 + a_2 \ln\left(Z_{2i}\right) + a_3 \ln\left(Z_{3i}\right) + \ldots + a_p \ln\left(Z_{pi}\right) + v_i \quad (6.15)$$

نحسب $LM = N * R^2$ من الانحدار المساعد ونقارنها بالقيمة الحرجة.

قدرنا الانحدار المساعد كما في المعادلة (6.15) وكانت النتائج كما يلي:

Dependent Variable: **LOG(U^2)**
Method: Least Squares
Date: 12/31/15 Time: 19:49
Sample: 1 20
Included observations: **20**

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	11.56297	4.219349	2.740464	0.0134
LOG(GDP)	0.270313	0.389046	0.694811	0.4960

R-squared	**0.026120**	Mean dependent var	14.45862
Adjusted R-squared	-0.027985	S.D. dependent var	2.907529
S.E. of regression	2.947932	Akaike info criterion	5.094724
Sum squared resid	156.4255	Schwarz criterion	5.194298
Log likelihood	-48.94724	Hannan-Quinn criter.	5.114162
F-statistic	0.482762	Durbin-Watson stat	1.645660
Prob(F-statistic)	0.496046		

ثم قدرنا احصائية $LM = N * R^2$ للانحدار المساعد، وكانت النتيجة كما يلي:

$$LM = 20 \times 0.026120 = 0.5224$$

ونقارن هذه القيمة بالقيمة الحرجة لكاي تربيع، وبما أن قيمة 0.5224 أصغر من قيمة χ^2 الحرجة، سنقبل الفرضية الأساسية القائلة بتجانس تباين حد الخطأ.

هـ) اختبار Goldfeld-Quandt test

اقترح Goldfeld and Quandt (1965) اختباراً بديلاً يقوم على فكرة أن تباين البواقي إذا كان ثابتاً لجميع المشاهدات (أي homoskedastic)، فإن تباين جزء من أجزاء العينة سيكون مساوياً لتباين جزء آخر من العينة، وحتى يكون الاختبار قابلاً للتطبيق، فمن الضروري تحديد المتغيّر المرتبط بتباين البواقي (يتم برسم البواقي بالنسبة للمتغيّرات التفسيرية)، ويتبع اختبار Goldfeld-Quandt test الخطوات التالية:

1- تحديد أحد المتغيّرات الذي يرتبط بتباين حد الخطأ وترتيب مشاهدات هذا المتغيّر تنازلياً (نبدأ من القيمة الأعلى ثم الأقل)

2- يتم تجزئة العينة المرتبة إلى جزئين متساويين بحذف المشاهدة المركزية c، وبالتالي تتكون العينة الجزئية من $\frac{1}{2}(n-c)$ مشاهدة، وتتضمن العينة الأولى القيم الكبيرة وتتكون العينة الثانية القيم الأدنى.

3- يتم تقدير انحدار المربعات الصغرى للمتغيّر التابع Y على المتغيّر المستقل X المستخدم في الخطوة (1) لكل عينة فرعية، ونحصل على مجموع SSR لكل معادلة.

4- احسب احصائية F كما يلي:

$$F = \frac{SSR_1}{SSR_2} \qquad (6.16)$$

يكون في البسط (SSR_1) للقيم الكبيرة، وتتوزع احصائية F بدرجات الحرية التالية: $F_{\left(\frac{1}{2}(n-c)-k,\ \frac{1}{2}(n-c)-k\right)}$

5- ترفض الفرضية الأساسية لتساوي التباين homoskedasticity إذا كانت $F_{الحرجة} <$ احصائية $-F$.

فالفكرة وراء هذه الصيغة هي أنه إذا كانت حدود الخطأ متساوية التباين homoskedastic سيكون تباين البواقي متساوٍ لكل عينة، وإذا كانت النسبة معنوية سترفض الفرضية الأساسية (تساوي التباين). (أنظر التطبيق في 6-4-2)

و) اختبار وايت White's test

طور White (1980) اختباراً أكثر عمومية لاختلاف التباين يزيل المشاكل التي تظهر في الاختبارات السابقة، وهو كذلك اختبار LM: إلا أن له المزايا التالية: (أ) لا يفترض المعرفة المسبقة عن اختلاف التباين، (ب) لا يعتمد على فرضية الطبيعية كاختبار Breusch-Pagan، (ج) يقترح خيار خاص لـ Z_S في الانحدار المساعد.

تفترض خطوات اختبار White نموذج بمتغيّرين تفسيريين كما يلي:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \qquad (6.17)$$

ويتبع نفس الخطوات في اختبارات LM باستثناء الخطوة الثانية التي هي تنفيذ الانحدار المساعد التالي:

$$\hat{u}_i^2 = a_1 + a_2 X_{2i} + a_3 X_{3i} + a_4 X_{2i}^2 + a_5 X_{3i}^2 + a_6 X_{2i} X_{3i} + v_i \quad (6.18)$$

أي انحدار مربع البواقي على: الحد الثابت (المقطع) وجميع المتغيّرات التفسيرية ومربع المتغيّرات التفسيرية وحاصل تقاطعهما.

ز) اختبار Engle's ARCH test

يستخدم هذا الاختبار على بيانات السلاسل الزمنية فقط، ويستخدم t كمؤشر للمتغيّرات، ويفحص هذا الاختبار وجود الارتباط الذاتي في حدود الخطأ لنموذج الانحدار. وقدم Engle (1982) مفهوم جديد لقبول الارتباط الذاتي لحدوثه في تباين حدود الخطأ بدلاً من حدود الخطأ نفسها. ولالتقاط الارتباط الذاتي هذا طور Engle نموذج الانحدار الذاتي المشروط بعدم ثبات تباين الأخطاء Autoregressive Conditional Heteroskedasticity (ARCH) والفكرة الأساسية له هي أن تباين u_t يعتمد على عدد فترات ابطاء مربع حدود الخطأ فترة واحدة $\left(u_{t-1}^2\right)$:

خذ نموذج الانحدار التالي لمزيد من التحليل:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \ldots + \beta_k X_{kt} + u_t \qquad (6.19)$$

افرض أن تباين حدود الخطأ تتبع عمليات $ARCH(1)$:

$$\text{var}(u_t) = \sigma_t^2 = \gamma_0 + \gamma_1 u_{t-1}^2 \qquad (6.20)$$

إذا لم يوجد ارتباط ذاتي في $\text{var}(u_t)$ ستكون γ_1 صفراً وبالتالي $\sigma_t^2 = \gamma_0$. وبالتالي ثبات التباين Homoskedasticity.

ويمكن توسيع النموذج لرتبة أعلى $ARCH(p)$:

$$\text{var}(u_t) = \sigma_t^2 = \gamma_0 + \gamma_1 u_{t-1}^2 + \gamma_2 u_{t-2}^2 + \cdots + +\gamma_p u_{t-p}^2 \qquad (6.21)$$

والفرضية الأساسية (عدم وجود أثر $ARCH$)

$$H_0 : \gamma_1 = \gamma_2 = \cdots = \gamma_p \quad (6.22)$$

تشمل خطوات اختبار أثر $ARCH$ ما يلي:

1- يتم تقدير نموذج (6.19) والحصول على بواقي $(\hat{u}_i)$.

2- تقدير انحدار مربع البواقي $(\hat{u}_t^2)$ على: الحد الثابت و $\hat{u}_{t-1}^2$ و $\hat{u}_{t-2}^2$ و و $\hat{u}_{t-p}^2$.

3- احسب احصائية $LM = (N-P)R^2$ من انحدار الخطوة (2)، فإذا كانت $(\text{LM}-\text{احصائية} > \chi_p^2)$ عند مستوى المعنوية المحدد ترفض الفرضية الأساسية (عدم وجود أثر ARCH) ونستنتج أن أثر ARCH موجود.

6-4- علاج اختلاف التباين

إذا وجد اختلاف التباين نستطيع حل المشكلة بطريقتين: (ا) نعيد تقدير النموذج بتطبيق طريقة المربعات الصغرى المعمّمَة generlized least squares (GLS) (أو المرجحة weighted least squares)، وهي تنتج مجموعة معلمات أكثر كفاءة من طريقة المربعات الصغرى العادية وتصحح مجموعة التباين المشترك وإحصائية t، أو (ب) وبما أن طريقة المربعات الصغرى العادية ليست الطريقة الفضلى مع أن نتائجها متسقة، إلا أن المشكلة الأساسية هي التباين المشترك وإحصائية t الخاطئة، ونستطيع تصحيح التباين المشترك وإحصائية t بالاستناد على الصيغة (6.8)، علماً بأنها لا تغيّر من قيمة المعلمات الفعلية المقدرة التي لا زالت أقل كفاءة.

6-4-1- طريقة المربعات الصغرى المعممة Generlized Least Squares (GLS)

خذ النموذج التالي

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_k X_{ki} + u_i \qquad (6.23)$$

نفترض أن تباين حد الخطأ غير متساوي، أي $\text{var}(u_i) = \sigma_i^2$ ، ولحل هذه المشكلة نقسم كل حد في المعادلة (6.24) على الانحراف المعياري لحد الخطأ σ_i ونحصل على نموذج معدّل:

$$\frac{Y_i}{\sigma_i} = \beta_1 \frac{1}{\sigma_i} + \beta_2 \frac{X_{2i}}{\sigma_i} + \beta_3 \frac{X_{3i}}{\sigma_i} + \cdots + \beta_k \frac{X_{ki}}{\sigma_i} + \frac{u_i}{\sigma_i} \qquad (6.24)$$

أو،

$$Y_i^* = \beta_1 X_{1i}^* + \beta_2 X_{2i}^* + \beta_3 X_{3i}^* + \cdots + \beta_k X_{ki}^* + u_i^* \qquad (6.25)$$

يكون تباين النموذج المعدل كما يلي:

$$\text{var}(u_i^*) = \text{var}\left(\frac{u_i}{\sigma_i^2}\right) = \frac{\text{var}(u_i)}{\sigma_i^2} = \frac{\sigma_i^2}{\sigma_i^2} = 1 \qquad (6.26)$$

وبالتالي يتم تقدير انحدار Y_i^* على X_{1i}^* و X_{2i}^* و X_{3i}^* و ... و X_{ki}^* باستخدام المربعات الصغرى العادية وبالتالي تصبح BLUE، ويسمى هذا الإجراء بالمربعات الصغرى المعمّمَة Generlized Least Squares (GLS).

مثال

إذا أردنا تقدير معادلة أسعار الشقق التي تأخذ الشكل التالي:

$$Price = \beta_1 + \beta_2\, Rooms + \beta_3\, Sqfeet + u_i$$

حيث يعتمد أسعار الشقق على عدد الغرف والمساحة، وتم تقدير المعادلة التالية:

Dependent Variable: PRICE
Method: Least Squares
Date: 12/31/15 Time: 20:52
Sample: 1 88
Included observations: 88

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-19315.00	31046.62	-0.622129	0.5355
ROOMS	15198.19	9483.517	1.602590	0.1127
SQFEET	128.4362	13.82446	9.290506	0.0000

R-squared	0.631918	Mean dependent var	293546.0
Adjusted R-squared	0.623258	S.D. dependent var	102713.4
S.E. of regression	63044.84	Akaike info criterion	24.97458
Sum squared resid	3.38E+11	Schwarz criterion	25.05903
Log likelihood	-1095.881	Hannan-Quinn criter.	25.00860
F-statistic	72.96353	Durbin-Watson stat	1.757956
Prob(F-statistic)	0.000000		

وتم إجراء اختبار Breusch-Pagan-Godfrey وكانت النتيجة كما يلي:

أظهر الاختبار أن إحصائية LM كانت 10.57632 وكانت معنوية عند مستوى معنوية يقل عن 1٪ ، أي يشير لوجود مشكلة Heteroskedasticity، وكانت النتائج كما يلي:

Heteroskedasticity Test: Breusch-Pagan-Godfrey

F-statistic	5.805633	Prob. F(2,85)	0.0043
Obs*R-squared	**10.57632**	**Prob. Chi-Square(2)**	**0.0051**
Scaled explained SS	23.13241	Prob. Chi-Square(2)	0.0000

Test Equation:
Dependent Variable: **RESID^2**
Method: Least Squares
Date: 12/31/15 Time: 20:54
Sample: 1 88
Included observations: **88**

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-8.22E+09	3.91E+09	-2.103344	0.0384
ROOMS	1.19E+09	1.19E+09	0.995771	0.3222
SQFEET	3881720.	1739736.	2.231213	0.0283

R-squared	**0.120185**	Mean dependent var	3.84E+09
Adjusted R-squared	0.099484	S.D. dependent var	8.36E+09
S.E. of regression	7.93E+09	Akaike info criterion	48.46019
Sum squared resid	5.35E+21	Schwarz criterion	48.54464
Log likelihood	-2129.248	Hannan-Quinn criter.	48.49421
F-statistic	5.805633	Durbin-Watson stat	1.649127
Prob(F-statistic)	0.004331		

وبما أن تباين حد الخطأ غير متساوٍ، أي $\text{var}(u_i)=\sigma_i^2$ ، ولحل هذه المشكلة نقسم كل حد في معادلة أسعار الشقق على الانحراف المعياري لحد الخطأ σ_i الذي يساوي 62315.97 كما يلي:

Price	Rooms	SqFeet	Price/ σ	Rooms/ σ	SqFeet/ σ
477500	7	3529	7.66256226	0.00011233	0.05663075
310000	6	1386	4.97464775	9.6284E-05	0.02224149
471250	5	2617	7.56226694	8.0236E-05	0.04199566
375000	5	2293	6.01771905	8.0236E-05	0.03679635
713500	5	3331	11.4497135	8.0236E-05	0.05345339
725000	5	3662	11.6342568	8.0236E-05	0.05876503
300000	5	2634	4.81417524	8.0236E-05	0.04226846
466275	5	2754	7.48243187	8.0236E-05	0.04419413
575000	5	3880	9.22716922	8.0236E-05	0.06226333
209000	4	1674	3.35387542	6.4189E-05	0.0268631
ملاحظة: هذه 10 مشاهدات من أصل العينة البالغة 88 مشاهدة					

ويعاد تقدير المعادلة لتصبح بعد التعديل كما يلي:

$$\frac{Price_i}{\sigma_i} = \beta_1 \frac{1}{\sigma_i} + \beta_2 \frac{Rooms_i}{\sigma_i} + \beta_3 \frac{Sqfeet_i}{\sigma_i} + \frac{u_i}{\sigma_i}$$

ونقدرها باستخدام المربعات الصغرى العادية؛ وتسمى هذا الطريقة بالمربعات الصغرى المعمّمة (GLS) Generlized Least Squares. وكانت نتيجة كما يلي:

Dependent Variable: PRICE
Included observations: 88
White heteroskedasticity-consistent standard errors & covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-19315.00	**41520.50**	-0.465192	**0.6430**
ROOMS	15198.19	**8943.735**	1.699311	**0.0929**
SQFEET	128.4362	**19.59089**	6.555914	**0.0000**

تم تصحيح الخطأ المعياري وحصلنا على تقدير أفضل (أكثر دقة).

6-4-2- طريقة المربعات الصغرى المرجحة

$$\mathrm{var}(u_i)=\sigma_i^2=\sigma^2 Z_i^2 \qquad (6.27)$$

حيث Z_i أحد المتغيّرات تكون جميع قيمه معروفة. وإجراء GLS هو نفسه المربعات الصغرى المرجحة weighted least squares (WLS)؛ حيث يكون لدينا الوزن ω الذي يصحح متغيّراتنا. لنرى تعريف $\omega_i=1/Z_i$ ونعيد كتابة النموذج الأصلي كما يلي:

$$\omega_i Y_i=\beta_1\omega_i+\beta_2(X_{2i}\,\omega_i)+\beta_3(X_{3i}\,\omega_i)+\ldots$$
$$+\beta_k(X_{ki}\,\omega_i)+(u_i\omega_i) \quad (6.28)$$

أو،

$$Y_i^*=\beta_1 X_{1i}^*+\beta_2 X_{2i}^*+\beta_3 X_{3i}^*+\cdots+\beta_k X_{ki}^*+u_i^* \qquad (6.29)$$

حيث تشير الحدود المنجمة (*) إلى متغيّرات مقسومة على Z_i، ويكون لدينا في هذه الحالة:

$$\mathrm{var}(u_i^*)=\mathrm{var}\left(\frac{u_i}{Z_i}\right)=\sigma^2 \qquad (6.30)$$

وبالتالي تحل مشكلة اختلاف التباين في النموذج الأصلي، مع ملاحظة أن هذه المعادلة ليس فيها حد ثابت، ويصبح الثابت في الانحدار الأصلي (β_1) معامل X_{1i}^* في (6.29)، وعلى فرض أن $Z_i=X_{3i}$، وبالتالي تصبح المعادلة كما يلي:

جدول (6-2) القيمة المضافة للصناعة التحويلية والناتج المحلي الاجمالي والسكان لعينة الدول العربية (القيمة بالمليون دولار)، 2010

الدولة	القيمة المضافة للصناعة MANU	الناتج المحلي الاجمالي GDP	عدد السكان POP	MANU/POP	GDP/POP
الأردن	4437	26970	6.111	726	4413
الامارات	28935	303812	8.264	3501	36763
البحرين	3923	20725	1.314	2986	15772
تونس	6602	42109	10.542	626	3994
الجزائر	8036	161736	35.847	224	4512
جيبوتي	30	1184	0.923	33	1283
السعودية	44757	457058	27.563	1624	16582
السودان	5904	69568	41.709	142	1668
سورية	2591	55621	20.618	126	2698
العراق	3300	123147	33.408	99	3686
عُمان	6170	60338	3.415	1807	17669
قطر	9403	122903	1.699	5534	72338
القمر	24	560	0.692	35	809
الكويت	6623	132065	3.582	1849	36869
لبنان	3007	39873	4.018	748	9924
ليبيا	3451	74763	7.774	444	9617
مصر	35166	225339	78.685	447	2864
المغرب	12909	96805	31.894	405	3035
موريتانيا	123	3701	3.570	34	1037
اليمن	2291	28181	23.154	99	1217

المصدر: صندوق النقد العربي، التقرير الإقتصادي العربي الموحد، 2011

$$\frac{Y_i}{Z_i}=\beta_1\frac{1}{Z_i}+\beta_2\frac{X_{2i}}{Z_i}+\beta_3\frac{X_{3i}}{Z_i}+\cdots+\beta_k\frac{X_{ki}}{Z_i}+\frac{u_i}{Z_i} \quad (6.31)$$

أو،

$$\frac{Y_i}{Z_i}=\beta_1\frac{1}{Z_i}+\beta_2\frac{X_{2i}}{Z_i}+\beta_3+\cdots+\beta_k\frac{X_{ki}}{Z_i}+\frac{u_i}{Z_i} \quad (6.32)$$

إذا استخدم هذا الشكل للمربعات الصغرى المرجحة *WLS*، سيتم الحصول على معاملات يجب أن تفسر بحذر شديد، وأصبحت المعلمة β_3 الآن الحد الثابت في (6.32)، بينما كانت معامل ميل في النموذج (6.23). ومن جهة أخرى، أصبحت المعلمة β_1 معامل ميل في (6.32) بينما كانت المقطع في النموذج الأصلي (6.23)، لذا يهتم الباحثون في أثر X_{3i} في (6.23) ويجب اختيار المقطع في (6.31) وبالمثل الحالات الأخرى.

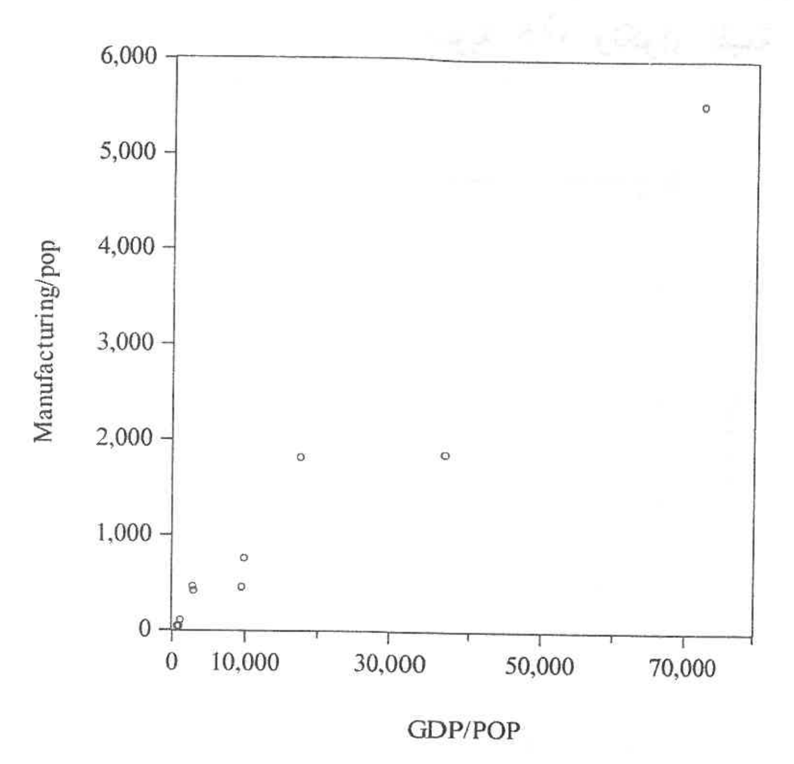
مثال

إذا أردنا تقدير انحدار القيمة المضافة في الصناعة (*MANU*) على الناتج المحلي الاجمالي (*GDP*) باستخدام بيانات الجدول (6-2) للنموذج التالي:

$$MANU=\beta_1+\beta_2 GDP+u_i \quad (6.33)$$

وعلى افتراض وجود مشكلة اختلاف التباين Heteroskedasticity، يكون أحد العلاجات قسمة جميع المشاهدات على عدد السكان وبالتالي يصبح النموذج كما يلي:

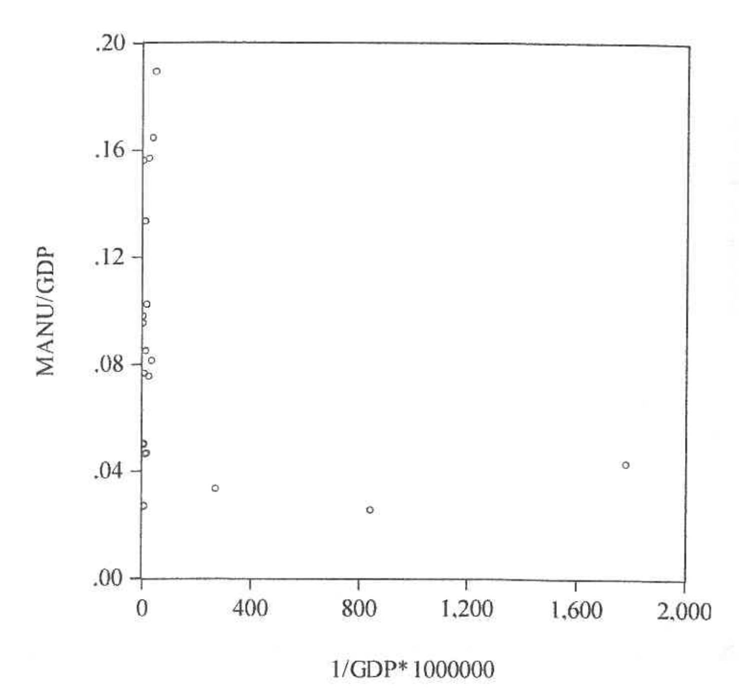
$$\frac{MA\hat{N}U}{POP} = \beta_1 \frac{1}{POP} + \beta_2 \frac{GDP}{POP} + \frac{u}{POP} \tag{6.34}$$



شكل رقم 6-5: حصة الفرد من التصنيع وحصته من الناتج المحلي الإجمالي

يبين الشكل (6-5) رسم $MANU/POP$ على GDP/POP، وبالرغم من القسمة على السكان يبدو الرسم يشبه Heteroskedasticity، وعندما نقدر (6.34) **نستخدم 10 دول حصة الفرد من الناتج المحلي الاجمالي فيها منخفضة و 10 دول مرتفعة الحصة، ومجموع مربعات البواقي** SSR **هي 1482436 للدول منخفضة الدخل و 1755079 للدول مرتفعة الدخل، ويتم قسمة القيمة الأولى على القيمة الثانية ونحصل على احصائية**

F التي تساوي **1.184**، وإذا كانت العينة الفرعية صغيرة، فمن الممكن الحصول على نسبة مرتفعة في ظل فرضية أساسية لاختلاف التباين، وفي هذه الحالة نرفضها عند مستوى معنوية 5٪، وتكون القيمة الحرجة لاحصائية $F_{(8,8)}$ هي 3.07.



شكل رقم 6-6: حصة الصناعة من الناتج المحلي الإجمالي ومعكوس الناتج المحلي الإجمالي

يظهر الشكل (6-6) نتائج رسم قيمة GDP على نفسه، وحصة GDP بالمقارنة مع معكوس GDP، وكان في هذه الحالة مجموع مربعات بواقي العينة الفرعية 21273566 و 340000000 وبالنهاية يكون لدينا نموذج ترفض فيه الفرضية الأساسية لثبات التباين.

سوف نقارن نتائج الانحدار الأصلي والنموذجين المحجمين، ملخصة في المعادلات التالية (الخطأ المعياري بين قوسين):

$$\hat{MANU} = \underset{(1426)}{-836} + \underset{(0.009)}{0.0999}\, GDP \quad , \quad R^2 = 0.86$$

$$\frac{\hat{MANU}}{POP} = \underset{(145)}{133}\,\frac{1}{POP} + \underset{(0.007)}{0.0763}\,\frac{GDP}{POP} \quad , \quad R^2 = 0.87$$

$$\frac{\hat{MANU}}{GDP} = \underset{(0.012)}{0.093} - \underset{(25.9)}{39}\,\frac{1}{GDP} \quad , \quad R^2 = 0.11$$

لاحظ ان تقدير معامل GDP هو نفسه تقريباً في الانحدارات الثلاثة، 0.099 و 0.076 و 0.093 (تذكر أنها تصبح المقطع بعد القسمة على المتغيّر GDP)، أحدها لا يتوقع انتقال دراماتيكي؛ حيث أن اختلاف التباين لا يرتفع إلى التحيّز، ومعلمات المعاملة الثالثة لها تباين أصغر وبالتالي يجب أن يكون الميل اكثر دقة، وربما يكون الخطأ المعياري أكبر، لكن الاخطاء المعيارية في أول انحدارين غير معتبرة لأنها غير صالحة بوجود اختلاف التباين.

لا يوجد تفسير اقتصادي للمقطع في هذا النموذج، وفي حالة تقديره في المعادلة الثالثة، حيث يصبح معامل $1/GDP$ غير معنوي ولا يختلف عن الصفر، ومشكلة النموذج هي أن قيمة R^2 منخفضة جداً.

6-4-3- النموذج غير الخطي

قد ينتج اختلاف التباين بسبب توصيف النموذج الرياضي، وعلى افتراض أن النموذج الرياضي الصحيح غير خطي هو:

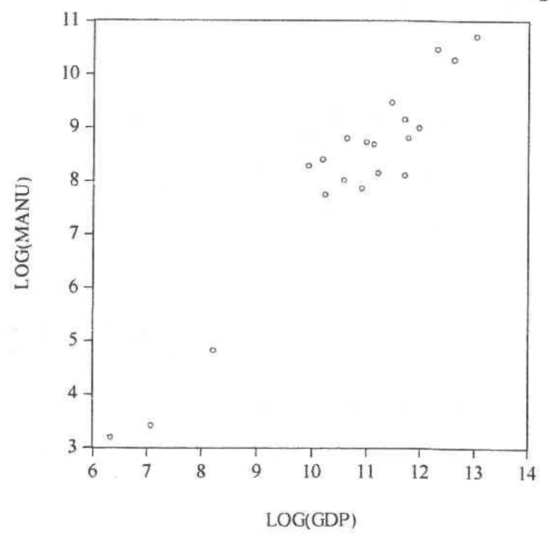
$$Y = \beta_1 X^{\beta_2} \upsilon \qquad (6.35)$$

بمعاملات موجبة β_1 و β_2 تكون Y دالة متزايدة في X، وضرب حد الخطأ υ له أثر في زيادة أو تخفيض Y بنسبة عشوائية على فرض أن توزيع υ الاحتمالي متساوي لجميع المشاهدات؛ فإن هذا يعني أن احتمالية 5٪ على سبيل المثال، يزيد أو يخفض Y نتيجة لأثرها نفسه عندما يكون X صغيراً، كما عندما يكون X كبيراً، وعلى كل حال، فإن الحد المطلق لزيادة 5٪ له تأثيراً أكبر على Y عندما تكون X أكبر منه عندما تكون Y صغيرة، وإذا رسمت Y مقابل X سيميل انتشار المشاهدات ليكون أكثر اتساعاً (بعثرة) حول العلاقة الصحيحة عندما يزيد X، وخط انحدار Y على X قد يبين اختلاف التباين.

الحل بتنفيذ لوغاريتم الانحدار:

$$\log Y = \beta_1 + \beta_2 \log X + u \qquad (6.36)$$

وهذا توصيف رياضي أكثر ملائمة، ويجعل نموذج الانحدار ثابت التباين، ويؤثر $\log\upsilon$ على المتغيّر التابع $\log Y$ بالاضافة إلى أن الحجم المطلق لأثرها هو استقلال $\log X$.



شكل رقم 6-7: لوغاريتم التصنيع ولوغاريتم الناتج المحلي الإجمالي

يظهر الشكل (6-7) لوغاريتم ناتج التصنيع مقابل لوغاريتم الناتج المحلي الاجمالي باستخدام بيانات جدول (6-2)، والنظرة الأولى للرسم لا تظهر اختلاف التباين، ويستخدم لوغاريتم الانحدار عينة فرعية لعشرة دول بأقل واكبر مجموع مربعات بواقي الناتج المحلي الاجمالي وهي 2.528104 و 2.337127 على التوالي، وبهذه الحالة وبما أن *SSR* الثانية أقل من الأولى سوف لا تكون معنوية كثيراً، وعلى كل حال يمكن استخدام اختبار Goldfeldt-Quandt test لاختبار اختلاف التباين، وحيث أن احصائية *F* تساوي 0.924 وهي أقل من قيمة F الحرجة عند مستوى معنوية 5٪، سوف لا نرفض الفرضية الأساسية لثبات التباين، وعند تقدير انحدار كامل العينة نحصل على:

$$\log \hat{MANU} = \underset{(0.832)}{-4.058} + \underset{(0.077)}{1.135} \log GDP, \qquad R^2 = 0.92 \qquad (6.37)$$

يعني أن مرونة القيمة المضافة للصناعة التحويلية *MANU* بالنسبة للناتج المحلي الاجمالي *GDP* تساوي 1 تقريباً.

لدينا نموذجين خاليين من اختلاف التباين (6.34) و (6.37)، ويخبرنا النموذج الأخير أن ناتج التصنيع يزداد تناسبياً مع الناتج المحلي الاجمالي لعينة مقطعية للدول العربية، ولنعمل خارج هذا التناسب نعيد كتابة المعادلة:

$$MANU = e^{-4.058} \quad GDP^{1.135} = 0.017\, GDP^{1.135} \qquad (6.38)$$

تخبرنا المعادلة (6.34) أن نسبة *MANU/GDP* أكثر فعالية ثابت، وبما أن الحد *1/GDP* يظهر أنه يتزايد، والثابت هو 0.01728.

6-4-4- طريقة تقدير اختلاف التباين المتسق Heteroskedasticity-consistent

اقترح (1980) White طريقة للحصول على تقدير متسق للتباين-التباين المشترك لمعلمات المربعات الصغرى، وسوف لا نعرض التفصيل الرياضي لهذه الطريقة هنا، ويستطيع EViews حساب White's Heteroskedasticity-corrected variance and standerd errors، وذلك بالنقر **Quick/Estimate Equation** وانقر على Options ثم انقر على صندوق Heteroskedasticity-consistent covariance ثم Next ثم OK.

تمارين

6-1- عرّف اختلاف التباين، واعطي مثالاً لنموذج قياسي يتضمن اختلاف التباين.

6-2- استخدم بيانات policy.wf1 لتقدير معادلة العلاقة بين القيمة الفعلية للموازنة الحالية Y القيمة المتوقعة للموازنة X، وافحص اختلاف التباين في معادلة الانحدار باستخدام جميع الاختبارات المعروفة والتي شرحت في هذا الفصل، وأعد تقدير نموذج تصحيح اختلاف التباين، وقارن النتائج التي حصلت عليها بنتائج تقدير الانحدار البسيط بطريقة المربعات الصغرى العادية.

6-3- بين كيفية تطبيق المربعات الصغرى المرجحة لحل مشكلة اختلاف التباين.

6-4- خذ النموذج التالي:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

حيث أن $\text{var}(u_i) = \sigma^2 X_{2i}$، جد تقدير المربعات الصغرى المعممّة.

6-5- صف اختبار Goldfeldt-Quandt test للكشف عن اختلاف التباين.

6-6- يتضمن الجدول بيانات الانفاق الحكومي G، والاستثمار I، والناتج المحلي الاجمالي Y، والسكان P لثلاثين دولة في عام

1997 (المصدر: صندوق النقد الدولي، الكتاب السنوي، 1997)، والقيم بالبليون دولار باستثناء عدد السكان بالمليون نسمة، وتحرى باحث فيما إذا كان الانفاق الحكومي يميل لمزاحمة الاستثمار وقدر الانحدار التالي (الخطأ المعياري بين قوسين):

$$\hat{I} = \underset{(7.79)}{18.10} - \underset{(0.14)}{1.07}\ G + \underset{(0.02)}{0.36}\ Y \qquad R^2 = 0.99$$

تم ترتيب المشاهدات تصاعدياً حسب حجم Y ونفذ الانحدار لـ 11 دولة الأقل دخلاً Y، و 11 دولة الأكبر دخلاً، وكان SSR للانحدارين 321 و 28101 على التوالي، وطبق اختبار Goldfeldt-Quandt test لبيان اختلاف التباين Heteroskedasticity.

Country	I	G	Y	P	Country	I	G	Y	P
Australia	94.5	75.5	407.9	18.5	Netherlands	73.0	49.9	360.5	15.6
Austria	46.0	39.2	206.0	8.1	New Zealand	12.9	9.9	65.1	3.8
Canada	119.3	125.1	631.2	30.3	Norway	35.3	30.9	153.4	4.4
Czech Republic	16.0	10.5	52.0	10.3	Philippines	20.1	10.7	82.2	78.5
Denmark	34.2	42.9	169.3	5.3	Poland	28.7	23.4	135.6	38.7
Finland	20.2	25.0	121.5	5.1	Portugal	25.6	19.9	102.1	9.8
France	255.9	347.2	1409.2	58.6	Russia	84.7	94.0	436.0	147.1
Germany	422.5	406.7	2102.7	82.1	Singapore	35.6	9.0	95.9	3.7
Greece	24.0	17.7	119.9	10.5	Spain	109.5	86.0	532.0	39.3
Iceland	1.4	1.5	7.5	0.3	Sweden	31.2	58.8	227.8	8.9
Ireland	14.3	10.1	73.2	3.7	Switzerland	50.2	38.7	256.0	7.1
Italy	190.8	189.7	1145.4	57.5	Thailand	48.1	15.0	153.9	60.6
Japan	1105.9	376.3	3901.3	126.1	Turkey	50.2	23.3	189.1	62.5
Korea	154.9	49.3	442.5	46.0	U.K.	210.1	230.7	1256.0	58.2
Malaysia	41.6	10.8	97.3	21.0	U.S.A.	1517.7	1244.1	8110.9	267.9

6-7- قدر الباحث في التمرين 6-6 أعلاه الانحدارات التالية

كمواصفات نموذج بديل (الخطأ المعياري بين قوسين):

$$\frac{\hat{I}}{P} = \underset{(0.28)}{-0.03}\frac{1}{P} - \underset{(0.16)}{0.69}\frac{G}{P} + \underset{(0.03)}{0.34}\frac{Y}{P} \quad , R^2 = 0.97 \quad (1)$$

$$\frac{\hat{I}}{Y} = \underset{(0.04)}{0.39} + \underset{(0.42)}{0.03}\frac{1}{Y} - \underset{(0.22)}{0.93}\frac{G}{Y} \quad , R^2 = 0.78 \quad (2)$$

$$\log\hat{I} = \underset{(0.26)}{-2.44} - \underset{(0.12)}{0.63}\log G + \underset{(0.12)}{1.60}\log Y, \quad R^2 = 0.98 \quad (3)$$

تم ترتيب العينة حسب Y/P و G/Y و $\log Y$ على التوالي، وقدر الانحدار في كل حالة مرة أخرى لعينات فرعية من مشاهدات 11 دولة من القيم الأصغر و 11 من القيم الأعظم. مبينة مجموع مربعات البواقي في الجدول:

	أصغر 11	أكبر 11
(1)	1.43	12.63
(2)	0.0223	0.0155
(3)	0.573	0.155

إجري اختبار Goldfeldt-Quandt test لمواصفات كل نموذج

ومناقشة مزايا كل المواصفات. وهل هناك أدلة على أن الاستثمار
هو دالة عكسية في الإنفاق الحكومي؟

الفصل السابع

الارتباط الذاتي

Autocorrelation

ينتهك الارتباط المتسلسل الفرضية (6) حيث تكون مشاهدات حد الخطأ المختلفة غير مرتبطة مع بعضها، ويسمى الارتباط المتسلسل Serial Correlation كذلك بالارتباط الذاتي Autocorrelation، ومن الممكن تواجده في أي دراسة بحثية، ويعني الارتباط المتسلسل أن قيمة حد الخطأ في أي فترة زمنية يعتمد على قيمة حد الخطأ في فترة أو فترات زمنية أخرى، وبما أن بيانات السلاسل الزمنية تستخدم في العديد من التطبيقات القياسية يكون واجباً علينا فهم الارتباط المتسلسل وعواقبه على مقدّرات OLS.

سنحاول في هذا الفصل الإجابة على نفس الأسئلة في الفصلين السابقين:

1- ما هي طبيعة المشكلة؟

2- ما هي نتائج المشكلة؟

3- كيف نشخص المشكلة؟

4- ما هي علاجات المشكلة المتاحة؟

7-1- طبيعة مشكلة الارتباط الذاتي

نعلم أن استخدام المربعات الصغرى العادية OLS لتقدير نموذج الانحدار يقودنا إلى تقدير BLUE للمعلمات، فقط عندما تكون جميع افتراضات نموذج الانحدار الخطي التقليدي متحققة، وفي هذا الفصل سنختبر أثر انتهاك فرضية عدم وجود ارتباط ذاتي، وهذه الفرضية نصت على يكون التباين المشترك والارتباط بين البواقي المختلفة مساوية جميعها للصفر:

$$cov(u_t, u_s) = 0 \qquad t \neq s \qquad (7.1)$$

بينت هذه الفرضية أن توزيع حدود الخطأ u_t و u_s مستقل، وتسمى بالتسلسل المستقل، فإذا لم تكن هذه الفرضية صحيحة فلا تكون أزواج حدود الخطأ مستقلة، ويكون بين هذه الأزواج ارتباطاً ذاتياً Autocorrelation (أو يوجد بينها ارتباط متسلسل Serially Correlated)، وفي هذه الحالة:

$$cov(u_t, u_s) \neq 0 \qquad t \neq s \qquad (7.2)$$

وهذا يعني أن الأخطاء عند الفترة t قد تكون مرتبطة بأحدها عند الفترة s.

غالباً ما يحدث الارتباط الذاتي في إطار سلسلة زمنية عندما تكون البيانات مرتبة ترتيباً زمنياً، وقد تؤثر الأخطاء في إحدى الفترات على الخطأ في فترة زمنية لاحقة (أو أخرى)، ومن المرجح أن يكون بين المشاهدات

المتعاقبة (أو المتتابعة) ارتباطاً بالأخص عندما تكون الفترات قصيرة مثل تكرار يومي أو أسبوعي أو شهري مقارنة مع بيانات مقطعية، مثلاً الزيادة غير المتوقعة في ثقة المستهلك قد تؤدي إلى تقدير معادلة استهلاك تكون أقل من تقدير الاستهلاك لفترتين أو أكثر، وقد نجد مشكلة الارتباط الذاتي في البيانات المقطعية، لكنها أقل احتمالاً؛ لأننا نستطيع بسهولة تغيير ترتيب البيانات بدون تغيير معنى النتائج.

7-1-1- أسباب حدوث الارتباط الذاتي

أحد العوامل التي قد تسبب الارتباط الذاتي هو حذف أو اسقاط متغيرات من النموذج، وعلى فرض أن Y_t مرتبط بالمتغيّرين X_{2t} و X_{3t}، إلا أننا بطريق الخطأ لم نضمّن المتغيّر X_{3t} في نموذجنا، سيتم التقاط تأثير المتغيّر X_{3t} بحد الخطأ u_t، فإذا كان X_{3t} سلسلة زمنية اقتصادية تعتمد على $X_{3,t-1}$ و $X_{3,t-2}$ وهكذا، سيؤدي هذا إلى ارتباط محتوم بين u_t و u_{t-1} و u_{t-2} وهكذا، وبالتالي قد يسبب اسقاط متغير إلى الارتباط الذاتي.

قد يحدث الارتباط الذاتي كذلك نتيجة سوء توصيف النموذج misspecification، وعلى فرض أن Y_t مرتبط بالمتغيّر X_{2t} بعلاقة تربيعية $Y_t = \beta_1 + \beta_2 X_{2t}^2 + u_t$، لكننا خطأً حددنا وقدرنا نموذج خطي $Y_t = \beta_1 + \beta_2 X_{2t} + u_t$ سنحصل على حدود خطأ لتوصيف خطي يعتمد على X_{2t}^2، فإذا زاد X_{2t} أو تناقص خلال الزمن ستتصرف u_t بنفس التصرف مشيرة إلى ارتباط ذاتي.

أما العامل الثالث هو الأخطاء المنهجية في القياس، وعلى فرض أن الشركة قامت بتحديث مخزونها في فترة زمنية معينة، وإذا حدثت أخطاء منهجية في قياسها سيُظهر المخزون التراكمي أخطاءً متراكمة في القياس، وستظهر هذه ارتباطاً ذاتياً.

7-1-2- الارتباط الذاتي من الدرجة الأولى ومن درجة أعلى

إن الحالة البسيطة والشائعة للارتباط الذاتي هي الارتباط المتسلسل serial correlation من الدرجة الأولى first order (الارتباط المتسلسل والارتباط الذاتي لهما نفس المعنى، وأي منهما يعني نفس المفهوم). افترض أن لديك نموذج الانحدار المتعدد التالي:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \cdots + \beta_k X_{kt} + u_t \qquad (7.3)$$

إن أي من مشاهدات حد الخطأ u_t الحالية هي دالة في المشاهدات السابقة (ابطائها lagged) لحد الخطأ u_{t-1}؛ أي أن:

$$u_t = \rho u_{t-1} + \varepsilon_t \qquad (7.4)$$

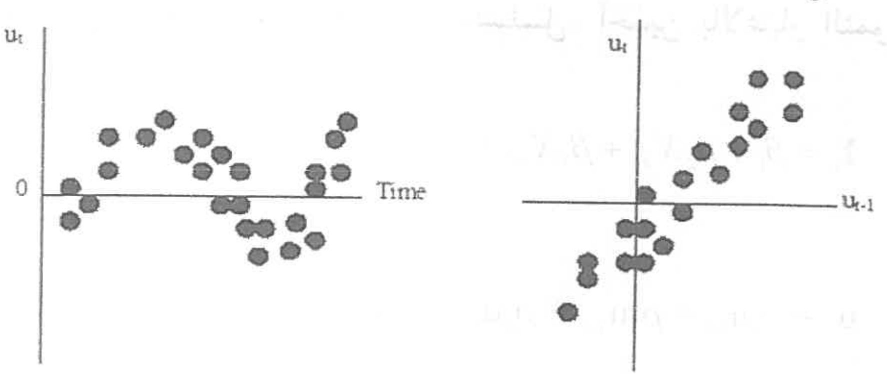
حيث تصور المعلمة ρ العلاقة الدالية بين مشاهدات حد الخطأ u_t، وأن الحد ε_t هو حد الخطأ الجديد، وتوزيعه متماثل ومستقل identically independently distributed (*iid*)، ويسمى المعامل ρ بمعامل الارتباط الذاتي من الدرجة الأولى، ويعطى قيمة تقع بين -1 و 1 أو لتجنب السلوك المتباعد explosive.

من الواضح أن حجم ρ يحدّد قوة الارتباط المتسلسل، ونستطيع التمييز بين ثلاث حالات من الارتباط المتسلسل:

أ- إذا كانت ρ تساوي صفراً، لا يوجد ارتباط متسلسل؛ لأن $u_t = \varepsilon_t$، وبالتالي تكون حدود الخطأ *iid*.

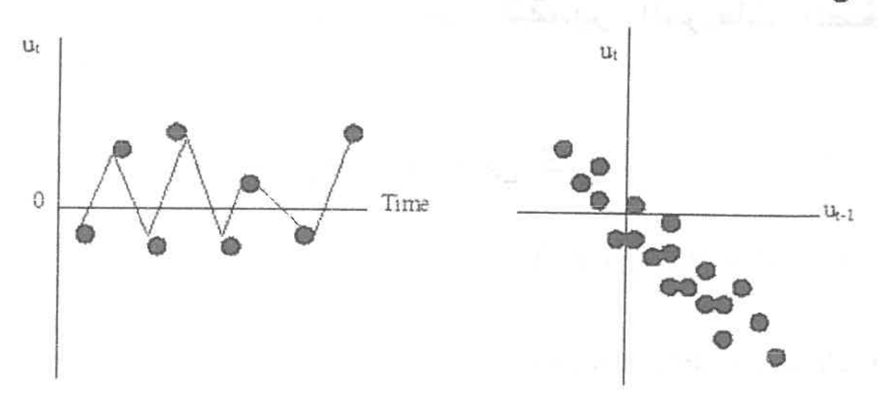
ب- إذا اقتربت قيمة ρ من +1 تصبح قيمة المشاهدة السابقة للخطأ u_{t-1} أكثر أهمية في تحديد قيمة حد الخطأ الحالي u_t، وبالتالي يوجد ارتباط متسلسل كبير موجب، وفي هذه الحالة تميل مشاهدات حد الخطأ الحالية

ليكون لها نفس إشارة مشاهدة حد الخطأ السابقة (الإشارة السالبة تقود إلى سالب، والموجبة تقود إلى موجب)، ويسمى هذا بالارتباط المتسلسل الموجب، ويبيّن الشكل (7-1) كيفية اظهار البواقي لحالة الارتباط الذاتي المتسلسل.



شكل رقم 7-1: الارتباط المتسلسل الموجب

ج- إذا كانت ρ تقترب من -1 ستكون قوة الارتباط الذاتي مرتفعة جداً، ويكون لدينا ارتباط متسلسل سالب؛ وهذا يعني أن السلوك يشبه أسنان المنشار عند رسم حدود الخطأ. وتميل اشارات حدود الخطأ للتحوّل من سالب إلى موجب والعكس صحيح في المشاهدات المتتالية، ويصوّر الشكل (7-2) حالة الارتباط المتسلسل السالب.



شكل رقم 7-2: الارتباط المتسلسل السالب

في الاقتصاد بشكل عام، يكون الارتباط المتسلسل السالب أقل حدوثاً من الارتباط المتسلسل الموجب.

وقد يأخذ الارتباط المتسلسل عدة أشكال، وقد يكون لدينا حدود خطأ تتبع درجات أعلى للارتباط المتسلسل، آخذين بالاعتبار النموذج التالي:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \cdots + \beta_k X_{kt} + u_t \qquad (7.5)$$

حيث أن:

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \rho_3 u_{t-3} + \ldots + \rho_P u_{t-P} + \varepsilon_t \qquad (7.6)$$

نقول في هذه الحالة لدينا P درجة ارتباط متسلسل، فإذا كان لدينا بيانات فصلية واسقط منها الأثر الموسمي مثلاً، ستتوقع وجود ارتباط متسلسل من الدرجة الرابعة، وبالمثل بيانات شهرية قد تظهر ارتباط متسلسل من الدرجة 12. وبشكل عام، فإن حالة ارتباط متسلسل من درجة أعلى لا يشبه حدوثه حدوث نوعٍ من الدرجة الأولى.

7-2- نتائج الارتباط الذاتي لتقدير المربعات الصغرى العادية

إذا كان لديك نموذج الانحدار الخطي التالي:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \cdots + \beta_k X_{kt} + u_t \qquad (7.7)$$

وإذا أظهر حد الخطأ u_t في هذه المعادلة ارتباطاً متسلسلاً، فإننا نستطيع تلخيص نتائج تقدير المربعات الصغرى العادية كما يلي:

1- لا زالت مقدرات المربعات الصغرى العادية OLS للمعلمات $\hat{\beta}$ غير منحازة ومتسقة، وهذا يسبب عدم التحيّز والاتساق، ولا يعتمد على الفرضية 6 لانتهاك الارتباط المتسلسل في هذه الحالة.

2- مقدرات OLS غير فعّالة وبالتالي لم تعد BLUE.

3- تباين معاملات الانحدار المُقدّر سيكون منحازاً وغير متسق، وبالتالي يصبح اختبار الفرضية غير صالح، وتكون قيمة R^2 (تشير إلى أفضل تقدير من تلك الموجودة) مرتفعة في اغلب الحالات، وتميل احصائية t لتكون مرتفعة (مشيرة إلى معنوية التقدير أكبر من الصحيح).

7-3- طرق اكتشاف الارتباط الذاتي

7-3-1- طريقة الرسم

إحدى أبسط الطرق لكشف الارتباط الذاتي هي الفحص برسم البواقي خلال الزمن ورسم انتشار u_t مقابل u_{t-1} وبيان فيما إذا أظهرت نمطاً من الأنماط التي عرضناها في الشكل (7-1) و (7-2)، وفي هذه الحالة يكون لدينا دليلاً عن ارتباط متسلسل موجب إذا كان النمط مشابهاً للشكل (7-1)، وارتباط متسلسل سالب إذا كان النمط مشابهاً للشكل (7-2).

مثال: اكتشاف الارتباط الذاتي باستخدام أسلوب الرسم

إذا أردنا اكتشاف الارتباط الذاتي باستخدام بيانات فصلية لتقدير المعادلة التالية:

$$C_t = b_1 + b_2 D_t + b_3 P_t + u_t$$

حيث أن:

C: انفاق المستهلكين على الطعام.

D: الدخل المتاح

P: الرقم القياسي للأسعار النسبية للطعام.

وتم تقدير المعادلة وكانت النتائج كما يلي:

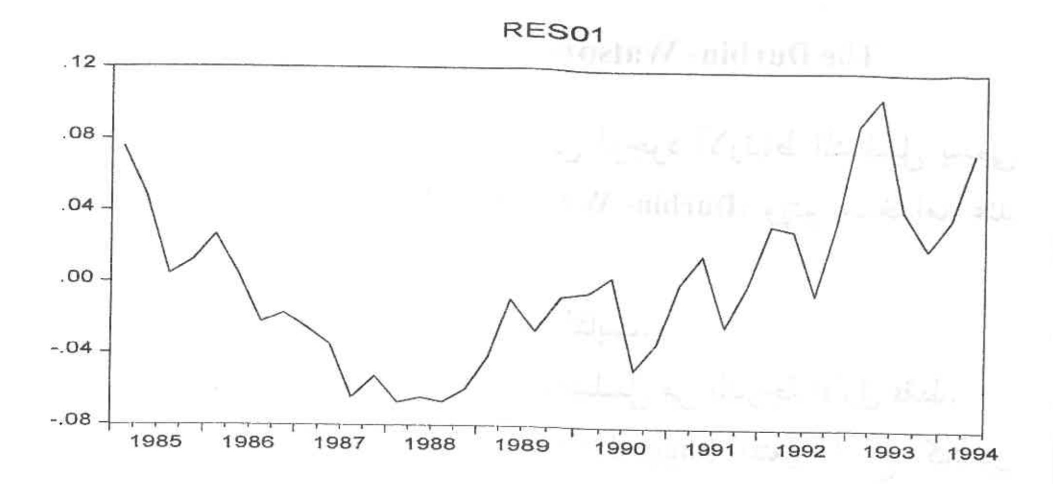
نتائج الانحدار

Dependent Variable: LCONS
Method: Least Squares
Date: 08/21/15 Time: 21:15
Sample: 1985Q1 1994Q2
Included observations: 38

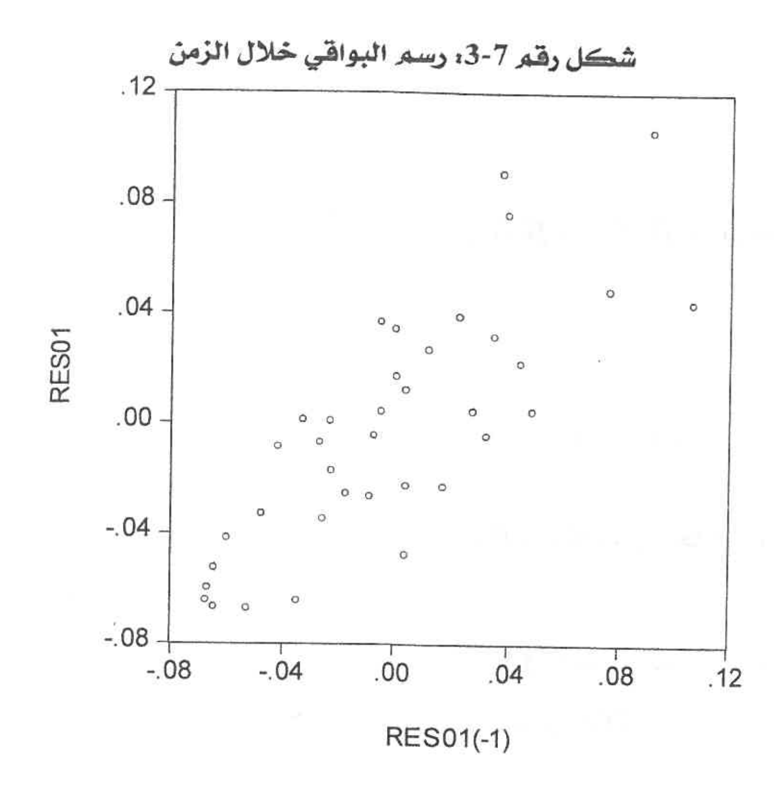
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.485434	0.788349	3.152708	0.0033
LDISP	0.529285	0.292327	1.810589	0.0788
LPRICE	-0.064029	0.146506	-0.437040	0.6648

R-squared	0.234408	Mean dependent var	4.609274
Adjusted R-squared	0.190660	S.D. dependent var	0.051415
S.E. of regression	0.046255	Akaike info criterion	-3.233656
Sum squared resid	0.074882	Schwarz criterion	-3.104373
Log likelihood	64.43946	Hannan-Quinn criter.	-3.187658
F-statistic	5.358118	Durbin-Watson stat	0.370186
Prob(F-statistic)	0.009332		

ولرسم البواقي نحسبها من خلال $resid = (Y-\hat{Y})$، ونرسمها مقابل الزمن كما في الشكل (7-3)، وكذلك نرسم الانتشار لها مقابل البواقي في الفترة *(t-1)* وتظهر في الشكل (7-4).



شكل رقم 7-3: رسم البواقي خلال الزمن



شكل رقم 7-4: رسم انتشار البواقي

يتضح من الشكل (7-3) و (7-4) أن البواقي لها ارتباط تسلسلي موجب.

7-3-2- اختبار دوربين- واتسون The Durbin- Watson test

غالباً ما يستخدم اختبار إحصائي لوجود الارتباط المتسلسل يسمى اختبار دوربين- واتسون Durbin- Watson (DW) test، ويتم استخدامه عند توفر الفرضيات التالية:

1- يتضمن نموذج الانحدار الحد الثابت.

2- يفترض أن يكون الارتباط المتسلسل من الدرجة الأولى فقط.

3- عدم تضمين المعادلة ابطاء (lagged) المتغيّر التابع كمتغير تفسيري.

فإذا كان لديك النموذج التالي:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \cdots + \beta_k X_{kt} + u_t \qquad (7.8)$$

حيث أن:

$$u_t = \rho u_{t-1} + \varepsilon_t \qquad |\rho| < 1 \qquad (7.9)$$

لاختبار الفرضية العدمية: $H_0 : \rho = 0$، ولتطبيق اختبار DW نتبع الخطوات التالية:

1- قدّر النموذج (7.8) باستخدام OLS واحسب البواقي $\hat{u}_t$.

2- احسب احصائية اختبار دوربين واتسون DW كما يلي:

$$d = \frac{\sum_{t=2}^{n} (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{T} \hat{u}_t^2} \qquad (7.10)$$

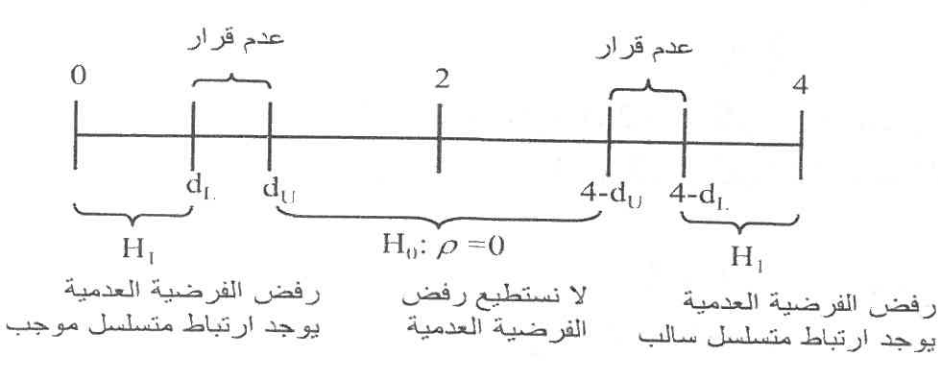
3- كوّن الشكل (7-5) معوضاً حساباتك d_U و d_L و $4-d_U$ و $4-d_L$ التي نحصل عليها من جدول القيم الحرجة لدوربين- واتسون (أنظر الملحق الاحصائي)، وجدول القيم الحرجة حسب k' التي هي عدد المتغيّرات التفسيرية بدون الحد الثابت.

4- اختبار الارتباط المتسلسل الموجب حسب الفرضية التالية:

لا يوجد ارتباط ذاتي $H_0 : \rho = 0$

يوجد ارتباط ذاتي موجب $H_1 : \rho > 0$

- إذا كانت $d \leq d_L$ نرفض H_0 ونستنتج وجود ارتباط متسلسل موجب.
- إذا كانت $d \geq d_U$ لا نستطيع رفض H_0 ونستنتج عدم وجود ارتباط متسلسل موجب.
- في حالة خاصة عندما $d_L < d < d_U$ يكون الاختبار غير حاسم.



شكل (7-5) اختبار دوربين- واتسون

5- لاختبار الارتباط المتسلسل السالب تكون الفرضية:

لا يوجد ارتباط ذاتي $H_0 : \rho = 0$

يوجد ارتباط ذاتي سالب $H_1 : \rho < 0$

- إذا كانت $d \geq 4 - d_L$ نرفض H_0 ونستنتج وجود ارتباط متسلسل سالب.
- إذا كانت $d \leq 4 - d_U$ لا نستطيع رفض H_0 ونستنتج عدم وجود ارتباط متسلسل سالب.
- في حالة خاصة عندما $4 - d_U < d < 4 - d_L$ يكون الاختبار غير حاسم.

وسبب عدم حسم اختبار دوربين- واتسون DW هو أن توزيع العينة الصغيرة لإحصائية DW يعتمد على المتغيرات X وصعوبة التحديد بشكل عام، ويفضل استخدام اختبار LM الذي سنوضحه لاحقاً.

قاعدة اختبار DW

من تقدير البواقي نستطيع الحصول على تقدير ρ كما يلي:

$$\hat{\rho} = \frac{\sum_{t=2}^{n} \hat{u}_t \hat{u}_{t-1}}{\sum_{t=1}^{n} \hat{u}_t^2} \qquad (7.11)$$

وتساوي احصائية DW تقريباً $d = 2\,(1 - \hat{\rho})$ لأن ρ لها مدى من -1 إلى 1، ومدى d يكون من 0 إلى 4، وبالتالي لدينا ثلاث حالات مختلفة:

أ- عندما $\rho = 0$ تكون $d = 0$، وبالتالي تكون قيمة d قريبة من 2 مشيرة إلى عدم وجود دليل على الارتباط المتسلسل.

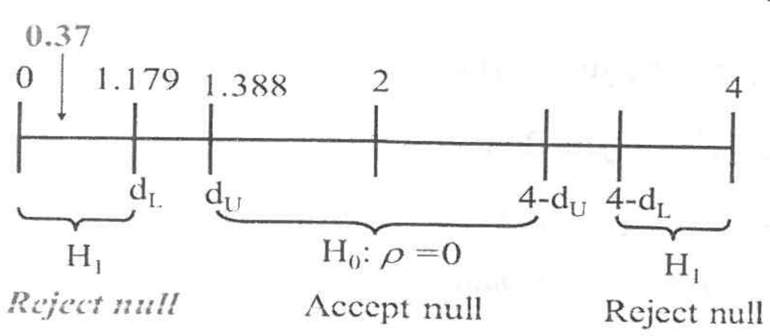
ب- إذا كانت $\rho \cong 1$ تكون $d \cong 0$ يكون ارتباطاً ذاتياً موجباً قوياً؛ وتعني أن ρ قريبة من +1، وبالتالي تكون قيمة d صغيرة جداً (قريبة من الصفر) للارتباط الذاتي الموجب.

ج- إذا كانت $\rho \cong -1$ تكون $d \cong 4$، وعندما تكون ρ قريبة من -1 تكون قيمة d قريبة من 4 مشيرة إلى ارتباط متسلسل سالب قوي.

ومن هذا التحليل نستطيع أن نرى أنه عندما تكون قيمة إحصائية اختبار DW قريبة جداً من 2 فهذا يعني أنه لا يوجد لدينا ارتباط متسلسل.

تطبيق

من نتائج الانحدار في المثال السابق نلاحظ أن إحصائية DW تساوي 0.37، ونجد أنه عند مستوى معنوية 1٪ والقيم الحرجة d_L و d_U لعينة $n = 38$ و $k' = 2$، تظهر النتائج في الشكل (7-6) حيث $d = 0.37$ أقل من $d_L = 1.176$، وبالتالي يكون لدينا دليلاً قوياً على ارتباط تسلسلي موجب.



شكل (7-6): مثال على اختبار DW

ملحوظة علمية

تزودنا البرمجيات بنتائج المربعات الصغرى متضمنة قيمة دوربن-واتسون بشكل تلقائي، إلا أن حساب قيمتها الحرجة أو قيمة p-value ليس سهلاً وأصبحت شعبيته كاختبار تتناقص، وكمؤشر لروه $rough(\rho)$ **فإذا كانت قيمة إحصائية دوربن-واتسون تساوي 1.4 أو أقل فهذا يعني وجود ارتباط ذاتي.**

7-3-3- اختبار Breusch-Godfrey LM test للارتباط المتسلسل

إن لاختبار DW العديد من العيوب التي تجعل من استخدامه غير مناسب في حالات مختلفة، مثلاً (أ) قد يعطي نتائج غير حاسمة، و (ب) غير قابل للتطبيق عند استخدام ابطاء المتغيّر التابع، و (ج) لا يمكن أن نأخذ بالاعتبار درجات أعلى للارتباط الذاتي.

لهذه الأسباب طوّر كل من *Breusch (1978)* و *Godfrey (1978)* اختبار *LM* الذي يستوعب جميع الحالات أعلاه، كما يلي:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \cdots + \beta_k X_{kt} + u_t \qquad (7.12)$$

حيث أن:

$$\mathrm{u_t} = \rho_1 \mathrm{u_{t-1}} + \rho_2 \mathrm{u_{t-2}} + \rho_3 \mathrm{u_{t-3}} + \ldots + \rho_\mathrm{P} \mathrm{u_{t-P}} + \varepsilon_\mathrm{t} \qquad (7.13)$$

مزج اختبار Breusch-Godfrey LM test هاتين المعادلتين كما يلي:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \cdots + \beta_k X_{kt} + \rho_1 \mathrm{u_{t-1}} + \rho_2 \mathrm{u_{t-2}} + \rho_3 \mathrm{u_{t-3}} + \ldots + \rho_\mathrm{P} \mathrm{u_{t-P}} + \varepsilon_\mathrm{t} \qquad (7.14)$$

وتكون الفرضية العدمية والبديلة كما يلي:

لا يوجد ارتباط ذاتي $H_0 : \rho_1 = \rho_2 = \ldots = \rho_\mathrm{P} = 0$

يوجد ارتباط ذاتي على الأقل أحد قيم ρ لا يساوي صفر : H_1

خطوات إجراء الاختبار

1- قدّر معادلة (7.12) باستخدام OLS للحصول على $\hat{u}_t$.

2- نفذ نموذج الانحدار التالي بعدة فترات إبطاء بافتراض درجة الارتباط الذاتي، مستخدمين P لتكون محدداً لدرجة الارتباط المتسلسل لاختبار:

$$\hat{u}_t = \alpha_0 + \alpha_1 X_{2t} + \alpha_2 X_{3t} + \cdots + \alpha_R X_{Rt} + \alpha_{R+1} u_{t-1} + \alpha_{R+2} u_{t-2} + \ldots + \alpha_{R+P} u_{t-P}$$

3- احسب احصائية LM التي تساوي $(n-p)R^2$ من انحدار الخطوة (2) أعلاه، فإذا كانت هذه الإحصائية أكبر من قيمة χ_P^2 الحرجة عند مستوى المعنوية، سنرفض الفرضية العدمية للارتباط المتسلسل ونستنتج وجود الارتباط المتسلسل، لاحظ أن اختيار P اعتباطي يعتمد على دورة البيانات: فصلية، شهرية، أسبوعية، يومية، ..

مثال: نستمر في علاقة الاستهلاك والدخل المتاح والأسعار، وسنستمر باختبار الارتباط المتسلسل من الدرجة 4 لأنه يوجد لدينا بيانات فصلية، ولإجراء الاختبار باستخدام Breusch-Godfrey LM test، نستخدم نتائج الانحدار ونجري اختبار بأربع فترات ابطاء ونحصل على:

نتائج اختبار Breusch-Godfrey LM test (من الدرجة الرابعة)

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	17.25931	Prob. F(4,31)	0.0000
Obs*R-squared	**26.22439**	**Prob. Chi-Square(4)**	**0.0000**

Test Equation:
Dependent Variable: RESID
Method: Least Squares
Date: 08/22/15 Time: 11:53
Sample: 1985Q1 1994Q2
Included observations: 38
Presample missing value lagged residuals set to zero.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.483704	0.489336	-0.988491	0.3306
LDISP	0.178048	0.185788	0.958341	0.3453
LPRICE	-0.071428	0.093945	-0.760322	0.4528
RESID(-1)	**0.840743**	**0.176658**	**4.759155**	**0.0000**
RESID(-2)	-0.340727	0.233486	-1.459306	0.1545
RESID(-3)	0.256762	0.231219	1.110471	0.2753
RESID(-4)	0.196959	0.186608	1.055465	0.2994

R-squared	0.690115	Mean dependent var	1.35E-17
Adjusted R-squared	0.630138	S.D. dependent var	0.044987
S.E. of regression	0.027359	Akaike info criterion	-4.194685
Sum squared resid	0.023205	Schwarz criterion	-3.893024
Log likelihood	86.69901	Hannan-Quinn criter.	-4.087356
F-statistic	11.50621	Durbin-Watson stat	1.554119
Prob(F-statistic)	0.000001		

ونرى من العمود الأول قيم كل من إحصائية LM وإحصائية F وقيمهما مرتفعة، وهذا يبين رفض الفرضية العدمية لعدم وجود ارتباط متسلسل، وكذلك يتضح من قيمة p-value انها صغيرة جداً (أقل من 0.05 لفترة ثقة 95٪)، وبالتالي يوجد ارتباط ذاتي، وإذا نظرنا إلى نتائج الانحدار نرى فقط أن الإبطاء الأول لحد الخطأ معنوي احصائياً، ومشيراً إلى أن الارتباط المتسلسل هو من الدرجة الأولى، ونعيد تنفيذ اختبار ارتباط متسلسل من الدرجة الأولى ونحصل على النتائج التالية:

نتائج اختبار Breusch-Godfrey LM test (من الدرجة الأولى)

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	53.47468	Prob. F(1,34)	0.0000
Obs*R-squared	**23.23001**	**Prob. Chi-Square(1)**	**0.0000**

Test Equation:
Dependent Variable: RESID
Method: Least Squares
Date: 08/22/15 Time: 12:21
Sample: 1985Q1 1994Q2
Included observations: 38
Presample missing value lagged residuals set to zero.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.585980	0.505065	-1.160208	0.2540
LDISP	0.245740	0.187940	1.307546	0.1998
LPRICE	-0.116819	0.094039	-1.242247	0.2226
RESID(-1)	**0.828094**	**0.113241**	**7.312638**	**0.0000**

R-squared	0.611316	Mean dependent var	1.35E-17
Adjusted R-squared	0.577020	S.D. dependent var	0.044987
S.E. of regression	0.029258	Akaike info criterion	-4.126013
Sum squared resid	0.029105	Schwarz criterion	-3.953636
Log likelihood	82.39425	Hannan-Quinn criter.	-4.064683
F-statistic	17.82489	Durbin-Watson stat	1.549850
Prob(F-statistic)	0.000000		

كذلك احصائية LM مرتفعة جداً كما هي احصائية t لإبطاء حد الخطأ، والارتباط الذاتي قطعاً من الدرجة الأولى.

7-3-4- اختبار Durbin's h test لإبطاء المتغير التابع

أشرنا سابقاً في فرضيات اختبار DW أنه غير قابل للتطبيق عندما يتضمن نموذج الانحدار إبطاءً للمتغيّر التابع كمتغيّر تفسيري، فإذا كان النموذج المراد اختباره يأخذ الشكل التالي:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \cdots + \beta_k X_{kt} + \gamma Y_{t-1} + u_t \qquad (7.14)$$

يكون اختبار DW غير صحيح.

ابتكر Durbin (1970) إحصائية اختبار يمكن استخدامها لمثل هذا النموذج، وتأخذ هذه الإحصائية (احصائية *h*) الشكل التالي:

$$h = \left(1 - \frac{DW}{2}\right)\sqrt{\frac{n}{1 - n\sigma_{\hat{\gamma}}^2}} \qquad (7.15)$$

حيث أن n عدد المشاهدات و DW احصائية دوربين- واتسون العادية المعرّفة بالمعادلة (7.10)، و $\sigma_{\hat{\gamma}}^2$ التباين المقدّر لمعامل إبطاء المتغيّر التابع، ولعينة كبيرة تتبع هذه الإحصائية التوزيع الطبيعي، وتشمل خطوات اختبار *h-test* ما يلي:

1- قدّر المعادلة (7.14) باستخدام OLS للحصول على البواقي، واحسب إحصائية DW حسب المعادلة (7.10).

2- احسب احصائية *h-statistics* حسب المعادلة (7.15).

3- الفرضية هي:

لا يوجد ارتباط ذاتي $H_0 : \rho = 0$

يوجد ارتباط ذاتي $H_1 : \rho \neq 0$

4- قارن احصائية h بالقيم الحرجة (في عينة كبيرة عند مستوى معنوية $\alpha = 0.05$ تكون القيمة الحرجة $z = \pm 1.96$)، فإذا تجاوزت احصائية h القيمة الحرجة سنرفض H_0 ونستنتج وجود ارتباط متسلسل.

مثال: اختبار دوربين h

إذا أردنا تقدير الانحدار التالي:

$$C_t = b_1 + b_2 D_t + b_3 P_t + b_4 C_{t-1} + u_t$$

الذي يتضمن إبطاء المتغيّر التابع، وفي هذه الحالة لا نستطيع استخدام اختبار DW، ونحتاج استخدام اختبار *Durbin's h test* أو *LM test*. وسنقدر أولاً الانحدار ونحصل على النتيجة التالية:

Dependent Variable: LCONS
Method: Least Squares
Date: 08/22/15 Time: 13:11
Sample (adjusted): 1985Q2 1994Q2
Included observations: 37 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.488356	0.575327	-0.848831	0.4021
LDISP	0.411340	0.169728	2.423524	0.0210
LPRICE	-0.120416	0.086416	-1.393442	0.1728
LCONS(-1)	0.818289	0.103707	7.890392	0.0000

R-squared	0.758453	Mean dependent var	4.608665
Adjusted R-squared	0.736494	S.D. dependent var	0.051985
S.E. of regression	0.026685	Akaike info criterion	-4.307599
Sum squared resid	0.023500	Schwarz criterion	-4.133446
Log likelihood	83.69058	Hannan-Quinn criter.	-4.246202
F-statistic	34.53976	**Durbin-Watson stat**	**1.727455**
Prob(F-statistic)	0.000000		

تساوي احصائية DW= 1.727455 ومن احصائية h نستطيع الحصول على القيمة التالية:

$$h = \left(1 - \frac{DW}{2}\right)\sqrt{\frac{n}{1 - n\sigma_{\hat{\gamma}}^2}}$$

$$= \left(1 - \frac{1.727455}{2}\right)\sqrt{\frac{37}{1 - 37 * 0.103707^2}} = 1.0689$$

حيث $\sigma_{\hat{\gamma}}^2$ هي تباين معامل إبطاء الاستهلاك $C = (0.103707)^2 = 0.0107551$، وعليه تكون احصائية h تساوي $h - statistics = 1.0689$، ولأنها أقل من القيمة الحرجة 1.96 نفشل في رفض الفرضية العدمية H_0 ونستنتج أن النموذج لا يعاني من ارتباط ذاتي.

وبتطبيق اختبار *LM test* لهذا الانحدار نحصل على النتيجة التالية:

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	0.680879	Prob. F(1,32)	0.4154
Obs*R-squared	**0.770865**	**Prob. Chi-Square(1)**	**0.3799**

ومن هذه النتائج يتضح عدم وجود ارتباط متسلسل في هذه النموذج.

7-4- علاج مشكلة الارتباط الذاتي

في حال وجود ارتباط ذاتي في النموذج سنحصل على مقدرات OLS غير كفؤة، ومن الضروري إيجاد طريقة لتصحيح تقديرنا، ويوجد لدينا حالتين مختلفتين لهما حل:

7-4-1- عندما تكون ρ معلومة

إذا كان لديك النموذج التالي:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \cdots + \beta_k X_{kt} + u_t \qquad (7.16)$$

حيث أن u_t لها ارتباط ذاتي ونتكهن بأنها تتبع ارتباط متسلسل من الدرجة الأولى، وعليه، فإن:

$$u_t = \rho u_{t-1} + \varepsilon_t \qquad (7.17)$$

إذا احتوت المعادلة (7.16) الفترة t، فإنها تحتوي كذلك على الفترة $t-1$:

$$Y_{t-1} = \beta_1 + \beta_2 X_{2t-1} + \beta_3 X_{3t-1} + \cdots + \beta_k X_{kt-1} + u_{t-1} \qquad (7.18)$$

نضرب جانبي المعادلة (7.18) بالمعلمة ρ ونحصل على:

$$\rho Y_{t-1} = \rho \beta_1 + \beta_2 \rho X_{2t-1} + \beta_3 \rho X_{3t-1} + \cdots + \beta_k \rho X_{kt-1} + \rho u_{t-1} \qquad (7.19)$$

نطرح المعادلة (7.19) من المعادلة (7.16) ونحصل:

$$Y_t - \rho Y_{t-1} = \beta_1(1-\rho) + \beta_2(X_{2t} - \rho X_{2t-1}) + \beta_3(X_{3t} - \rho X_{3t-1}) + \cdots + \beta_k(X_{kt} - \rho X_{kt-1}) + (u_t - \rho u_{t-1}) \qquad (7.20)$$

أو،

$$Y_t^* = \beta_1^* + \beta_2 X_{2t}^* + \beta_3 X_{3t}^* + \cdots + \beta_k X_{kt}^* + \varepsilon_t \qquad (7.21)$$

حيث أن $Y_t^* = (Y_t - \rho Y_{t-1})$ ، و $\beta_1^* = \beta_1(1-\rho)$ ، و $X_{it}^* = (X_{it} - \rho X_{it-1})$.

عند أخذ الفرق سنخسر إحدى المشاهدات، ولتجنب هذه الخسارة يقترح تحويل Y_1 و X_1 للمشاهدة الأولى كما يلي:

$$Y_1^* = Y_1\sqrt{1-\rho^2} \quad , \quad X_{i1}^* = X_{i1}\sqrt{1-\rho^2} \tag{7.22}$$

تسمى طريقة التحويل هذه للمتغيرات حيث أن Y_t^* ، و β_1^* ، و X_{it}^* بشبه الفرق أو الفرق المعمم، لاحظ أن حد خطأ المعادلة (7.21) يحقق جميع فرضيات نموذج الانحدار الخطي التقليدية، فإذا كانت ρ معروفة نستطيع تطبيق OLS على المعادلة (7.21) ونحصل على تقدير BLUE، والمثال الذي يستخدم طريقة الفرق المعمم يظهر أدناه:

نتائج الانحدار الذي يحدد قيمة ρ

Dependent Variable: RES01
Method: Least Squares
Date: 08/22/15 Time: 15:31
Sample (adjusted): 1985Q2 1994Q2
Included observations: 37 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
RES01(-1)	**0.799544**	0.100105	7.987073	0.0000

R-squared	0.638443	Mean dependent var	-0.002048
Adjusted R-squared	0.638443	S.D. dependent var	0.043775
S.E. of regression	0.026322	Akaike info criterion	-4.410184
Sum squared resid	0.024942	Schwarz criterion	-4.366646
Log likelihood	82.58841	Hannan-Quinn criter.	-4.394835
Durbin-Watson stat	1.629360		

نتائج انحدار الفرق المعمّم

Dependent Variable: LCONS-STAR
Method: Least Squares
Date: 08/22/15 Time: 15:42
Sample (adjusted): 1985Q2 1994Q2
Included observations: 38

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C_STAT	4.089403	1.055839	3.873131	0.0004
LDISP_STAR	0.349452	0.231708	1.508155	0.1405
LPRICE_STAR	-0.235900	0.074854	-3.151460	0.0033
R-squared	0.993284	Mean dependent var		0.974724
Adjusted R-squared	0.992900	S.D. dependent var		0.302420
S.E. of regression	0.025482	Akaike info criterion		-4.426070
Sum squared resid	0.022726	Schwarz criterion		-4.296787
Log likelihood	87.09532	**Durbin-Watson stat**		**1.686825**

7-4-2- عندما تكون ρ مجهولة

يوجد طريقة سهلة تشبه طريقة التحويل، حيث أن ρ مجهولة، سنحتاج لإجراء يزودنا بتقدير ρ ثم تقدير نموذج انحدار (7.21)، طورت عدة إجراءات من أشهرها وأكثرها أهمية هي: (أ) إجراء تكرار Vochrane-Orcutt iterative procedure، و (ب) إجراء بحث Hildreyh-Lu search procedure وسنتعرف عليهما فيما يلي:

أ- إجراء تكرارات Vochrane-Orcutt iterative procedure

طور Vochrane-Orcutt (1949) إجراء التكرارات الذي سنعرضه بالخطوات التالية:

1- قدّر نموذج انحدار المعادلة (7.16) ومنها نحصل على البواقي u_t.

2- قدر معامل ρ للارتباط المتسلسل من الدرجة الأولى باستخدام OLS لـ $\hat{u}_t = \rho \hat{u}_{t-1} + \varepsilon_t$.

3- حوّل المتغيّرات الأصلية لتصبح حيث أن $Y_t^* = (Y_t - \hat{\rho} Y_{t-1})$، و $\beta_1^* = \beta_1(1-\hat{\rho})$، و $X_{it}^* = (X_{it} - \hat{\rho} X_{it-1})$، و $Y_1^* = Y_1\sqrt{1-\hat{\rho}^2}$ و $X_{i1}^* = X_{i1}\sqrt{1-\hat{\rho}^2}$.

4- نفذ الانحدار مستخدماً المتغيّرات المحوّلة، وجد البواقي لهذا الانحدار، وبما أننا لا نعلم فيما إذا كانت $\hat{\rho}$ الناتجة من الخطوة 2 تقدير مناسب للمعلمة ρ أم لا، ارجع إلى الخطوة 2 واعد الخطوات من 2-4 عدة مرات إلى أن تصل إلى معادلة خالية من الارتباط المتسلسل.

ب- إجراء Hildreyh-Lu search procedure

طور Hildreyh-Lu (1960) أسلوب بديل لإجراء -Vochrane Orcutt، ويتكون هذا الأسلوب من الخطوات التالية:

1- اختر قيمة ρ (مثل ρ_1) وحول قيم النموذج كما في (7.21)، وقدره باستخدام OLS.

2- نحصل على بواقي $\hat{\varepsilon}_t$ من نتائج تقدير الخطوة 1 وعلى مجموع مربع بواقي $SSR_{(\rho_1)}$، ثم اختر قيم مختلفة للمعلمة ρ (مثل ρ_2) واعد الخطوة 1 و 2.

3- بتنويع ρ من -1 إلى +1 بطريقة منهجية نستطيع الحصول على قيم $SSR_{(\rho_i)}$، ونحتاج ρ يكون مجموع مربعات البواقي SSR لها الأدنى minimized، وتكون المعادلة (9.21) التي قدرناها باستخدام ρ المختارة الحل المثل.

هذا الإجراء معقّد ويشتمل عدة حسابات.

7-5- مثال كامل لاختبار الارتباط الذاتي

لتقدير دالة الادخار في الأردن تم وصف معادلة الادخار كما يلي:

$$S = \alpha + \psi\, GDP + \theta\, r$$

حيث أن:

S : الادخار الإجمالي المتاح في السنة t

GDP : الناتج المحلي الإجمالي في السنة t

r : أسعار الفائدة على الودائع طويلة الأجل في السنة t

1- تم استخدام بيانات الجدول (8-1) **لتقدير المعادلة**. وكانت النتائج كما يلي:

$$\underset{t=}{S} = \underset{1.457}{606.247} + \underset{7.983}{0.180}\, GDP - \underset{-1.295}{61.053}\, r$$

$$R^2 = 0.851$$

$$\text{DW} = 1.012427$$

2- اختبار الارتباط الذاتي:

أ- غالباً ما يستخدم اختبار **دوربين- واتسون Durbin- Watson (DW) test**، للكشف عن وجود الارتباط المتسلسل، واختبار الفرضية العدمية التالية:

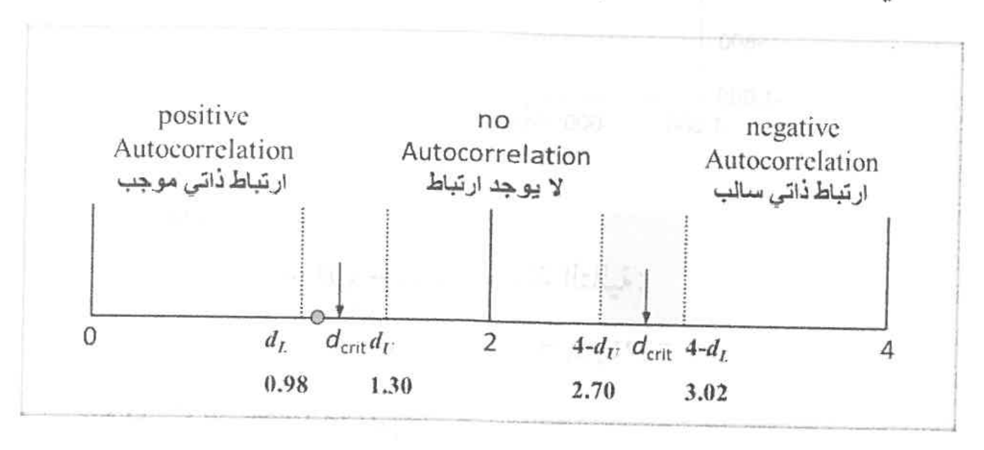
جدول (8-1) الناتج المحلي الإجمالي، والإدخار، وأسعار الفائدة

السنة	الناتج المحلي الإجمالي GDP	الادخار المتاح الإجمالي S	اسعار الفائدة على الودائع طويلة الأجل r
1985	1970.5	253.9	8.5
1986	2240.5	380.1	7.5
1987	2286.7	318.5	7.5
1988	2349.5	418.9	8.0
1989	2425.4	531.5	8.0
1990	2760.9	318.2	8.2
1991	2958.0	424.8	7.8
1992	3610.5	611.0	7.0
1993	3884.2	956.3	6.9
1994	4357.4	1143.2	7.3
1995	4714.7	1374.4	8.0
1996	4911.3	1341.9	8.9
1997	5137.4	1342.3	8.9
1998	5609.9	1239.5	8.3
1999	5778.1	1533.3	7.9
2000	5998.6	1360.8	6.6
2001	6363.7	1322.2	5.2
2002	6794.0	1720.9	4.0
2003	7228.8	2356.2	2.8
2004	8090.7	2243.2	2.5
2005	8925.4	1437.4	3.5
2006	10675.4	1801.5	5.1
2007	12131.4	1633.9	5.6
2008	15593.4	3216.4	5.7
2009	16912.2	3648.9	4.2

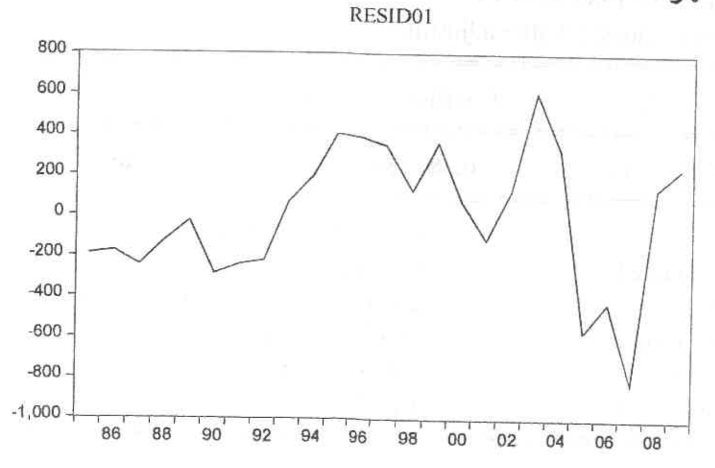
لا يوجد ارتباط ذاتي $H_0 : \rho = 0$

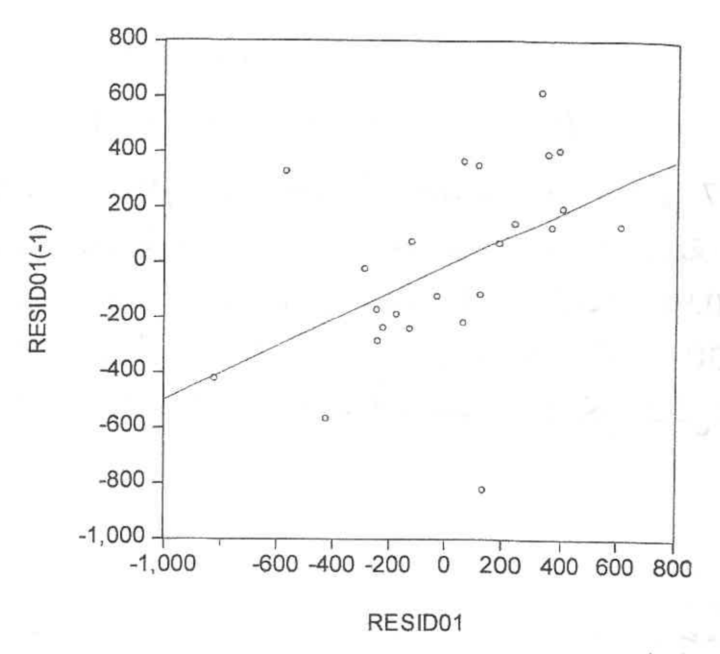
يوجد ارتباط ذاتي موجب $H_1 : \rho > 0$

نلاحظ من نتائج هذا الانحدار أن إحصائية DW تساوي 1.012427، وعند مستوى معنوية 1٪ تكون القيم الحرجة d_L و d_U لعينة $n = 25$ و $k' = 2$ أي عدد المعلمتين بدون معلمة الحد الثابت $d_L = 0.98$ و $d_U = 1.30$، وبما أن $d = 1.012$ تقع بين $d_L = 0.98$ و $d_U = 1.30$، وبالتالي يكون لدينا حالة عدم الحسم كما تظهر في الشكل التالي:



ب- كما تُظهر نتائج رسم البواقي مع الزمن، ومع إبطائها فترة زمنية واحدة، تظهر هذه المشكلة بجلاء:





3- الحل

أ- سنقدر قيمة ρ ، حسب المعادلة التالية:

$$u_t = \rho u_{t-1} + \varepsilon_t \qquad |\rho| < 1$$

وكانت النتيجة كما يلي:

Dependent Variable: RESID01
Method: Least Squares
Date: 01/05/16 Time: 19:27
Sample (adjusted): 1986 2009
Included observations: 24 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
RESID01(-1)	**0.486950**	0.183140	**2.658892**	**0.0140**

R-squared	0.234687	Mean dependent var	7.896671
Adjusted R-squared	0.234687	S.D. dependent var	342.7375
S.E. of regression	299.8342	Akaike info criterion	14.28511
Sum squared resid	2067713.	Schwarz criterion	14.33420
Log likelihood	-170.4213	Hannan-Quinn criter.	14.29813
Durbin-Watson stat	1.804723		

وكانت هذه النتيجة تؤكد معنوية ρ.

أو باستخدام الصيغة التالية: $d = 2\,(1-\hat{\rho})$ ومنها نستخرج قيمة ρ

$$d = 2\,(1-\hat{\rho})$$
$$1.012427 = 2-2\hat{\rho}$$
$$1.012427 - 2 = -2\hat{\rho}$$
$$-0.987573 = -2\hat{\rho}$$
$$\hat{\rho} = 0.4937865$$

وكانت هذه النتيجة لا تختلف عن التقدير في المعادلة أعلاه.

ب- سنحول بيانات المتغيّرات الأصلية لتصبح كما يلي:

$$Y_t^* = \left(Y_t - \hat{\rho}\, Y_{t-1}\right)$$

وإذا أخذنا قيمة $\hat{\rho} = 0.486950$ سنحول البيانات كما يلي:

$$S_2^* = \left(380.1 - 0.486950(\,253.9)\right) = 256.4634$$

$$S_3^* = \left(318.5 - 0.486950(\,380.1)\right) = 133.4103$$

$$\vdots$$

وهكذا.

- ثمّ نحوّل بيانات $X_{it}^* = \left(X_{it} - \hat{\rho}\, X_{it-1}\right)$ ، وإذا أردنا تحويل بيانات GDP تصبح كما يلي:

$$GDP_2^* = \left(2240.5 - 0.486950(\,1970.5)\right) = 1280.965$$

$$GDP_3^* = \left(2286.7 - 0.486950(\,2240.5)\right) = 1195.689$$

$$\vdots$$

ونستمر هكذا.

- ثم نحول بيانات سعر الفائدة r بنفس الطريقة.

- ثم نحصل على القيمة الأولى من كل متغيّر كما يلي:

$$S_1^* = S_1\sqrt{1-\hat{\rho}^2}$$

$$S_1^* = 253.9\,\sqrt{1-0.486950^2} = 221.764$$

ثمّ

$$r_1^* = r_1\sqrt{1-\hat{\rho}^2}$$

$$r_1^* = 8.5\,\sqrt{1-0.486950^2} = 7.424$$

ثمّ

$$GDP_1^* = GDP_1\sqrt{1-\hat{\rho}^2}$$

$$GDP_1^* = 1970.5\ \sqrt{1-0.486950^2} = 1721.094$$

وهذا ما نشاهده في الجدول (8-2) التالي:

جدول (8-2) الناتج المحلي الإجمالي، والإدخار، وأسعار الفائدة والقيم المحوّلة لها

Year	GDP	S	r	Sstar	GDPstar	rstar
1985	1970.5	253.9	8.5	**221.764**	**1721.094**	**7.424**
1986	2240.5	380.1	7.5	256.463395	1280.965025	3.360925
1987	2286.7	318.5	7.5	133.410305	1195.688525	3.847875
1988	2349.5	418.9	8.0	263.806425	1235.991435	4.347875
1989	2425.4	531.5	8.0	327.516645	1281.310975	4.1044
1990	2760.9	318.2	8.2	59.386075	1579.85147	4.3344
1991	2958.0	424.8	7.8	269.85251	1613.579745	3.8124015
1992	3610.5	611.0	7.0	404.14364	2170.1019	3.142051
1993	3884.2	956.3	6.9	658.77355	2126.067025	3.4856975
1994	4357.4	1143.2	7.3	677.529715	2465.98881	3.9846535
1995	4714.7	1374.4	8.0	817.71876	2592.86407	4.4006565
1996	4911.3	1341.9	8.9	672.63592	2615.476835	4.9690085
1997	5137.4	1342.3	8.9	688.861795	2745.842465	4.6004925
1998	5609.9	1239.5	8.3	585.867015	3108.24307	3.9912755
1999	5778.1	1533.3	7.9	929.725475	3046.359195	3.8337065
2000	5998.6	1360.8	6.6	614.159565	3184.954205	2.7079645
2001	6363.7	1322.2	5.2	659.55844	3442.68173	2.0004775
2002	6794.0	1720.9	4.0	1077.05471	3695.196285	1.4427295
2003	7228.8	2356.2	2.8	1518.207745	3920.4617	0.8168085
2004	8090.7	2243.2	2.5	1095.84841	4570.63584	1.1508875
2005	8925.4	1437.4	3.5	345.07376	4985.633635	2.3074945
2006	10675.4	1801.5	5.1	1101.55807	6329.17647	3.415936
2007	12131.4	1633.9	5.6	756.659575	6933.01397	3.0619465
2008	15593.4	3216.4	5.7	2420.772395	9686.01477	2.952558
2009	16912.2	3648.9	4.2	2082.67402	9318.99387	1.473863

ج- ثمّ نعيد تقدير المعادلة حسب البيانات المحوّلة ونحصل على النتيجة التالية:

$$\text{Sstar} \underset{t=}{=} \underset{1.118}{277.099} + \underset{6.329}{0.191}\,\text{GDPstar} - \underset{-1.295}{57.428}\,rstar$$

$$R^2 = 0.741$$

$$\text{DW} = 1.793744$$

أو،

Dependent Variable: SSTAR
Method: Least Squares
Date: 01/05/16 Time: 21:08
Sample: 1985 2009
Included observations: 25

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	277.0993	247.8262	1.118120	0.2756
GDPSTAR	0.191042	0.030182	6.329730	0.0000
RSTAR	-57.42895	50.38659	-1.139766	0.2666

R-squared	0.741509	Mean dependent var	745.5609
Adjusted R-squared	0.718009	S.D. dependent var	574.8138
S.E. of regression	305.2422	Akaike info criterion	14.39225
Sum squared resid	2049801.	Schwarz criterion	14.53852
Log likelihood	-176.9032	Hannan-Quinn criter.	14.43282
F-statistic	31.55461	**Durbin-Watson stat**	**1.793744**
Prob(F-statistic)	0.000000		

ثم نختبر قيمة إحصائية **Durbin- Watson (DW)** وهي تقدر بحوالي 1.8 وهي قريبة جداً من القيمة 2، وبالتالي فهي تشير إلى عدم وجود ارتباط ذاتي في هذه المعادلة وبذلك تم حل هذه المشكلة.

ولمزيد من التأكد نجري اختبار Breusch-Godfrey LM test كما يلي:

1- بعد تقدّير المعادلة المعدّلة باستخدام OLS تم الحصول على البواقي لها $\hat{u}_t$.

2- سنقدر نموذج الانحدار التالي بفترة إبطاء واحدة:

$$\hat{S}star = \alpha_0 + \alpha_1 GDPstar + \alpha_2 rstar + \alpha_3 u_{t-1}$$

وكانت النتيجة كما يلي:

Dependent Variable: **RESID02**
Method: Least Squares
Date: 01/05/16 Time: 21:35
Sample (adjusted): 1986 2009
Included observations: 24 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-22.72434	310.4689	-0.073194	0.9424
GDPSTAR	0.003895	0.033833	0.115126	0.9095
RSTAR	2.365949	68.66821	0.034455	0.9729
RESID02(-1)	0.107961	0.238139	0.453350	0.6552

R-squared	0.010710	Mean dependent var	-1.758988
Adjusted R-squared	-0.137683	S.D. dependent var	298.3975
S.E. of regression	318.2774	Akaike info criterion	14.51474
Sum squared resid	2026010.	Schwarz criterion	14.71108
Log likelihood	-170.1768	Hannan-Quinn criter.	14.56683
F-statistic	0.072176	Durbin-Watson stat	1.986022
Prob(F-statistic)	0.974191		

3- نحسب احصائية LM من انحدار الخطوة (2) أعلاه كما يلي:

$$\begin{aligned} LM &= (n-p)R^2 \\ &= (25)(0.010710) \\ &= 0.26775 \end{aligned}$$

ونقارن هذه الإحصائية بالقيمة الحرجة لقيمة $\chi_P^2 = 37.65$ عند مستوى المعنوية 5٪، وبما أن قيمة LM أقل من القيمة الحرجة 37.65 سنقبل الفرضية العدمية للارتباط المتسلسل ونستنتج بعدم وجوده.

وهذه النتيجة كانت مطابقة للاختبار الجاهز في برمجية EViews التالية:

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	0.236453	Prob. F(1,21)	0.6318
Obs*R-squared	0.278357	Prob. Chi-Square(1)	0.5978

Test Equation:
Dependent Variable: RESID
Method: Least Squares
Date: 01/05/16 Time: 21:26
Sample: 1985 2009
Included observations: 25
Presample missing value lagged residuals set to zero.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-37.65101	263.8586	-0.142694	0.8879
GDPSTAR	0.004643	0.032169	0.144329	0.8866
RSTAR	6.476631	52.98576	0.122233	0.9039
RESID(-1)	0.111625	0.229557	0.486264	0.6318

R-squared	0.011134	Mean dependent var	-8.64E-14
Adjusted R-squared	-0.130132	S.D. dependent var	292.2471
S.E. of regression	310.6811	Akaike info criterion	14.46106
Sum squared resid	2026978.	Schwarz criterion	14.65608
Log likelihood	-176.7632	Hannan-Quinn criter.	14.51515
F-statistic	0.078818	Durbin-Watson stat	1.991729
Prob(F-statistic)	0.970780		

تمارين

7-1- عرّف المصطلحات التالية:

أ) الارتباط المتسلسل. وارتباط متسلسل من الدرجة الأولى.

ب) معامل الارتباط الذاتي من الدرجة الأولى.

ج) إحصائية دوربين واتسون.

د) ارتباط متسلسل موجب.

7-2- استخدم الجدول (ب) في الملحق لاختبار الارتباط المتسلسل لاحصائية دوربين واتسون التالية:

أ) d=0.81، و k=3 و N=21 عند مستوى معنوية 5٪.

ب) d=3.48، و k=2 و N=15 عند مستوى معنوية 5٪.

ج) d=1.56، و k=5 و N=30 عند مستوى معنوية 5٪.

د) d=2.84، و k=4 و N=35 عند مستوى معنوية 5٪.

هـ) d=1.75، و k=1 و N=45 عند مستوى معنوية 5٪.

7-3- لديك بيانات الاستهلاك الخاص LCONS، والدخل المتاح LDISP، ومستوى الأسعار LPRICE، وأردت دراسة دالة الاستهلاك لهذه الدولة (الخطأ المعياري بين قوسين):

$$LCÔNS = 2.48 - \underset{(0.29)}{0.52} \ LDISP - \underset{(0.15)}{0.06} \ LPRICE$$

$$R^2 = 0.23 \qquad DW = 0.37$$

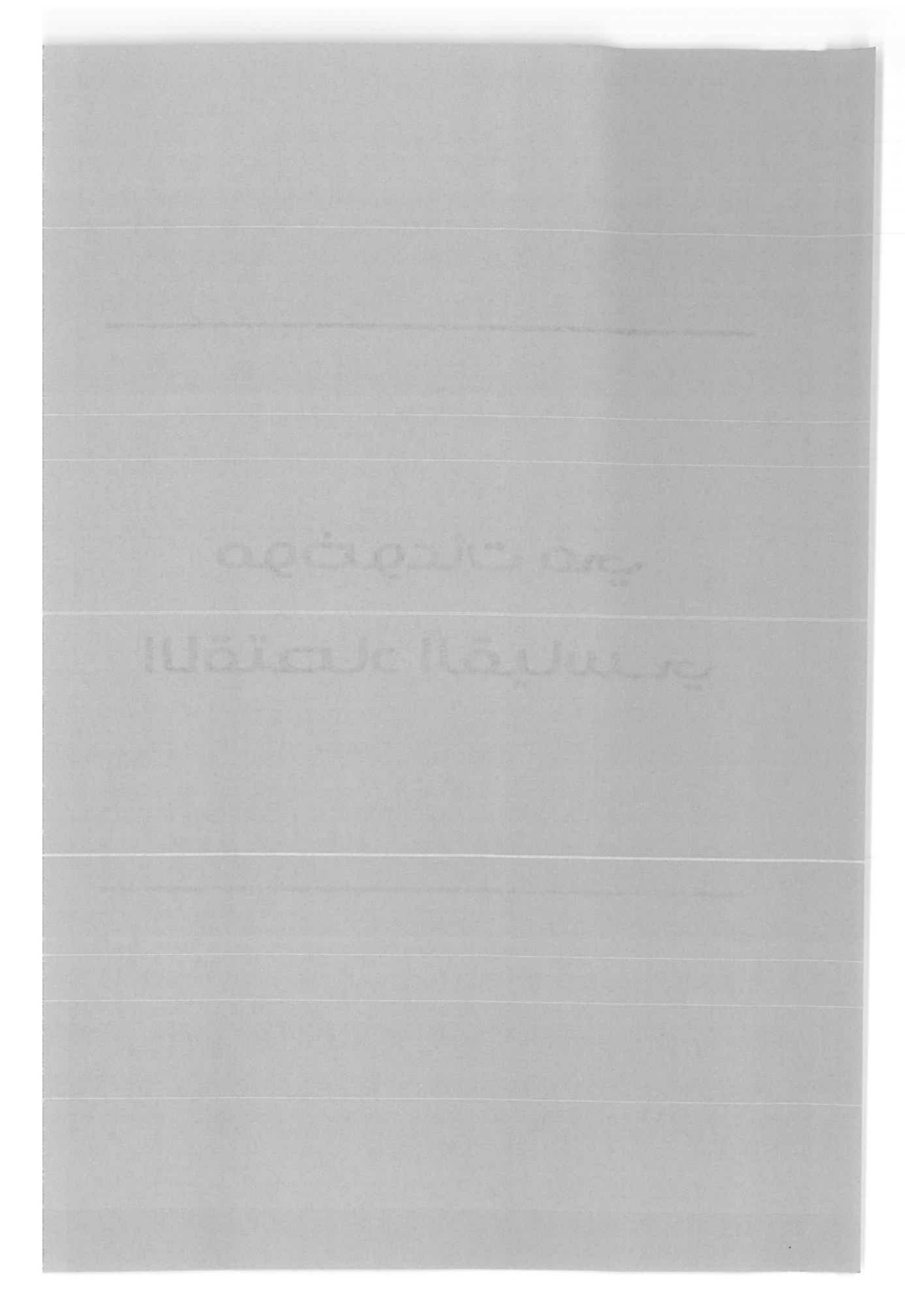
أ) بين فيما إذا كانت هذه المعادلة تعاني من مشكلة الارتباط الذاتي أم لا، مستخدماً اختبار DW عند مستوى معنوية 5٪.

ب) ما هي المشكلة القياسية إن وجدت في هذه المعادلة؟ ماذا تقترح لحل مشكلة الارتباط الذاتي إن وجد في هذه المعادلة.

ج) إفرض أنك أضفت إبطاء المتغير التابع بفترة واحدة إلى هذه المعادلة، فهل تطبق هنا اختبار دوربين واتسون أم لا؟ لماذا، وإذا لا فلماذا لا؟

موضوعات في

الاقتصاد القياسي



الفصل التاسع

المتغيّرات الوهمية

Dummy Variables

من الممكن أن تكون بعض المتغيّرات المستخدمة في نموذج الانحدار نوعية qualitative وغير كمية؛ ومن الأمثلة على ذلك:

1- قد تريد التحقق من وجود علاقة بين الدراسة والأجر المكتسب، وقد يتضمن هذا المثال متغيّراً يتضمن كل من الذكور والإناث، وترغب بمعرفة وجود فروقات في الأجر حسب الجنس.

2- قد تريد التحقق من وجود علاقة بين الدخل والإنفاق في الأردن لعينة تتضمن عائلات أردنية وعائلات غير أردنية (سورية، عراقية، مصرية مثلاً) وترغب بمعرفة فيما إذا كان اختلاف الجنسية يحقق فروقات في الاستهلاك.

3- إذا كان لديك بيانات عن معدل نمو حصة الفرد من الناتج المحلي الإجمالي وحصة الفرد من المساعدات الأجنبية لعينة دول نامية بعضها ديموقراطية والأخرى غير ذلك، وترغب من التحقق فيما إذا كان أثر المساعدات الأجنبية على النمو يتأثر بنوعية الحكومة.

كل مثال من هذه الأمثلة يتم حله عن طريق تقدير انحدارين منفصلين لكل فئة، وسترى فيما إذا كان معامل كل منهما مختلفاً عن الآخر أم لا، أما الحل البديل الآخر هو تقدير انحدار واحد مستخدماً جميع المشاهدات لقياس أثر العامل النوعي المسمى "**المتغيّر الوهمي Dummy variable**"، وهذا الحل له ميزتين مهمتين: أولاً الحصول على أسلوب بسيط لاختبار فيما إذا أثر العامل النوعي مؤثراً أم لا، كما يبين لنا فيما إذا الافتراض صحيحاً، وجاعلاً تقدير الانحدار أكثر كفاءة.

8-1- استخدام المتغيّر الوهمي

إذا أردنا دراسة تكاليف المدارس الثانوية وأنها تختلف باختلاف نوع المدرسة فيما إذا كانت مدارس أكاديمية أم مدارس مهنية، سنستخدم المتغيّرات الوهمية Dummy Variables للتمييز بين تكاليف هذين النوعين من المدارس. في البداية دعنا نقدّر في البداية معادلة تكلفة المدارس الثانوية، حيث أنها تختلف باختلاف عدد الطلاب بغض النظر عن نوعية المدرسة، وسنبدأ بالنموذج التالي:

$$COST = \beta_1 + \beta_2 N + u \qquad (8.1)$$

حيث أن $COST$ النفقات السنوية المترتبة على المدرسة، و N عدد الطلاب، وبيّنت النتائج التالية انحدار عينة 74 مدرسة ثانوية في شنغهاي في منتصف الثمانينات، وكانت كما يلي:

Dependent Variable: **COST**
Method: Least Squares
Date: 08/18/15 Time: 17:35
Sample: 1 74
Included observations: 74

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	23953.30	27167.96	0.881674	0.3809
N	339.0432	49.55144	6.842248	0.0000

R-squared	0.394023	Mean dependent var	187418.0
Adjusted R-squared	0.385606	S.D. dependent var	141969.9
S.E. of regression	111280.6	Akaike info criterion	26.10415
Sum squared resid	**8.92E+11**	Schwarz criterion	26.16642
Log likelihood	-963.8536	Hannan-Quinn criter.	26.12899
F-statistic	46.81636	Durbin-Watson stat	1.352470
Prob(F-statistic)	0.000000		

كانت معادلة الانحدار التي حصلنا عليها كما يلي (الانحراف المعياري بين قوسين):

$$\hat{COST} = \underset{(27000)}{24000} + \underset{(50)}{339}N \qquad R^2 = 0.39 \qquad (8.2)$$

تم قياس الكلفة بالإيوان Yuan (كان 1 يوان يساوي 20 سنت أمريكي وقت إجراء المسح)، وتعني هذه المعادلة أن الكلفة الحدية لكل طالب هي 339 يوان، وكلفة النفقات السنوية العامة (الإدارية والصيانة) هي 24000 يوان.

هذه هي نقطة البداية لنرى المعادلة العامة لتكاليف المدارس الثانوية، ثم نريد التحقق من أثر نوعية المدرسة (المدارس العادية والمدارس المهنية) على الكلفة؛ حيث تهدف المدارس المهنية إلى إعطاء مهارات مهن معينة ويكون تشغيلها مكلف نسبياً لأنها تحتاج إلى ورش عمل متخصصة، ولحل هذه المشكلة سيتم تقدير معادلتين منفصلتين لكل نوع من أنواع المدارس، ويكون لدينا معادلتين كما يلي:

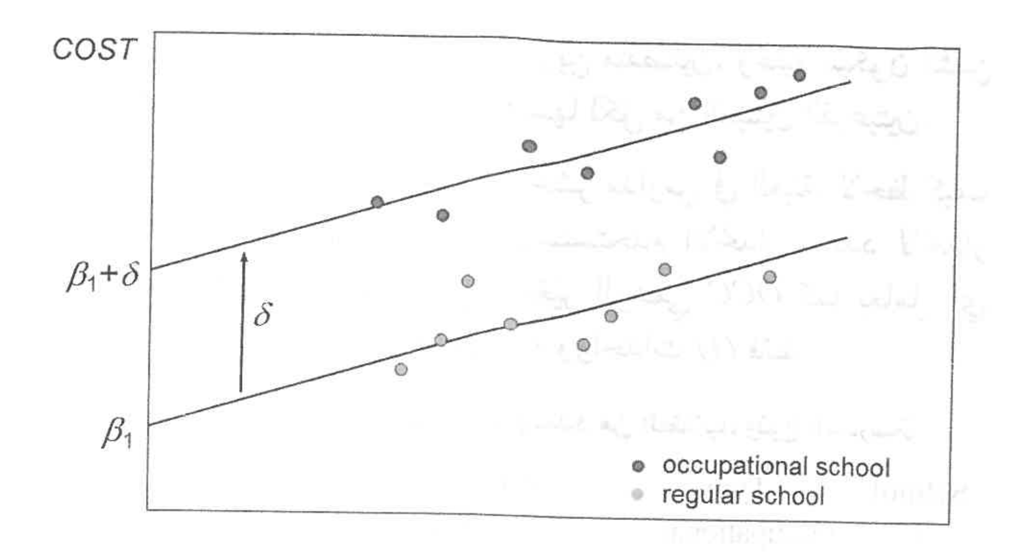
$$COST = \beta_1 + \beta_2 N + u \qquad (8.3)$$

$$COST = \beta_1' + \beta_2 N + u \qquad (8.4)$$

تتعلق المعادلة الأولى بالمدارس العادية والثانية بالمدارس المهنية، ومن الطبيعي افتراض أن النفقات الكلية السنوية تختلف حسب نوع المدرسة، وأن الكلفة الحدية هي نفسها لكليهما، وهذا الافتراض للكلفة الحدية غير معقول، ويمكن التخفيف منه في الوقت المناسب. دعنا نعرّف δ لتكون الفرق بين المقطعين $\delta = \beta_1' - \beta_1$، وبالتالي فإن $\beta_1' = \beta_1 + \delta$، وتستطيع إعادة كتابة دالة تكاليف المدارس المهنية كما يلي:

$$COST = \beta_1 + \delta + \beta_2 N + u \qquad (8.5)$$

ويبين الشكل (8-1) رسم هذا النموذج، حيث يظهر الخطين العلاقة بين الكلفة وعدد الطلاب، مهملين حد الخطأ، وخط المدارس المهنية هو نفس خط المدارس العادية ما عدا انتقاله إلى الأعلى بمقدار δ.



شكل رقم 8-1: دالة تكاليف المدارس المهنية والمدارس العادية

الهدف من هذا التمرين هو تقدير عامل الانتقال غير المعروف، ولإجراء هذا نعيد كتابة النموذج كما يلي:

$$COST = \beta_1 + \delta\, OCC + \beta_2 N + u \qquad (8.6)$$

حيث أن *OCC* متغيّر وهمي؛ وهو متغيّر مصطنع يتضمّن قيمتين محتملتين هما *0* و *1*؛ حيث تم تكوينه كمتغيّر جديد لتمييز نوع المدرسة؛ فإذا كانت مدرسة اكاديمية عادية يكون رمزها *0*، أما إذا كانت مهنية يكون رمزها *1*، وبذلك يصبح لدينا متغيّر مكوّن من *0* و *1* كما هو في الجدول *(8-1)*. وبعد تقدير المعادلة، فإذا كان *OCC* يساوي *0* تكون المعادلة هي معادلة الكلفة *(8.3)* للمدارس الأكاديمية العادية، أما إذا كانت تساوي *1* تكون معادلة دالة الكلفة *(8.5)* للمدارس المهنية، وبدلاً من تقدير انحدارين منفصلين سنستخدم العينة كاملة في تقدير انحدار واحد يختصر تباينات مجتمع المعلمات التي تُعكس بأخطاء معيارية أقل، وسنحصل كذلك على

تقدير واحد للمعلمة β_2 بدلاً من تقديرين منفصلين، وعليه سيكون الثمن الذي ندفعه هو افتراض أن β_2 هي نفسها لكل من العينتين الفرعيتين.

يبين الجدول *(8-1)* بيانات أول عشر مدارس في العينة، لاحظ كيف يختلف *OCC* حسب نوع المدرسة، وسنستخدم الانحدار المتعدد لانحدار *COST* على *N* و *OCC*، ويعامل المتغيّر الوهمي *OCC* كما يعامل أي متغيّر عادي، وهو يتكون من أصفار *(0)* وواحدات *(1)* فقط.

جدول (8-1) النفقات المتكررة، وعدد من الطلاب، ونوع المدرسة

School	Type	COST	N	OCC
1	Occupational	345,000	623	1
2	Occupational	537,000	653	1
3	Regular	170,000	400	0
4	Occupational	526	663	1
5	Regular	100,000	563	0
6	Regular	28,000	236	0
7	Regular	160,000	307	0
8	Occupational	45,000	173	1
9	Occupational	120,000	146	1
10	Occupational	61,000	99	1

تعطينا نتائج تقدير المعادلة (8.6) نتائج انحدار لعينة الكاملة (74 مدرسة)، وكانت النتائج كما يلي (الخطأ المعياري بين قوسين):

$$\hat{COST} = \underset{(24000)}{-34000} + \underset{(21000)}{133000}\ OCC + \underset{(40)}{331}N \qquad R^2 = 0.62 \qquad (8.7)$$

فإذا جعلنا *OC7C* تساوي *0* و *1* على التوالي، نستطيع الحصول على دوال الكلفة الضمنية لكل من نوعي المدارس:

المدارس العادية: $\hat{C}OST = -34000 + 331N$

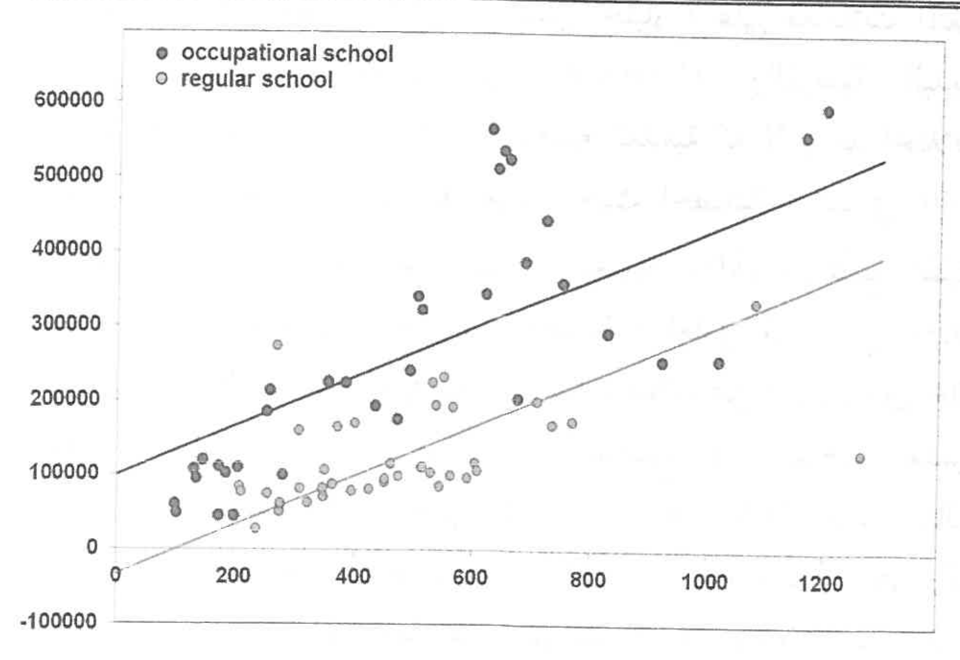
المدارس المهنية : $\hat{C}OST = -34000 + 133000 + 331N$

$$= 99000 + 331N$$

Dependent Variable: COST
Method: Least Squares
Date: 08/18/15 Time: 21:26
Sample: 1 74
Included observations: 74

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-33612.55	23573.47	-1.425864	0.1583
N	331.4493	39.75844	8.336578	0.0000
OCC	133259.1	20827.59	6.398201	0.0000

R-squared	0.615637	Mean dependent var	187418.0
Adjusted R-squared	0.604810	S.D. dependent var	141969.9
S.E. of regression	89248.09	Akaike info criterion	25.67592
Sum squared resid	**5.66E+11**	Schwarz criterion	25.76933
Log likelihood	-947.0092	Hannan-Quinn criter.	25.71319
F-statistic	56.86072	Durbin-Watson stat	1.299150
Prob(F-statistic)	0.000000		



شكل رقم 8 -2: دالة تكاليف المدارس العادية والمهنية في شنغهاي

يعني الانحدار أن الكلفة الحدية لكل طالب في السنة هي 331 يوان والكلفة السنوية العامة للمدارس الأكاديمية العادية هي - 34000 يوان، ومن الواضح لدينا مقطع سالب ليس له أي معنى بشكل عام، ويبين أن النموذج فيه خطأ توصيف، ومعلمة المتغيّر الوهمي 133000 هي تقدير للتكاليف العامة السنوية للمدارس المهنية، والكلفة الحدية للمدارس المهنية هي نفسها للمدارس العادية، ويبين الشكل (8-2) البيانات ودالة الكلفة المشتقة من نتائج الانحدار.

8-1-1- الخطأ المعياري واختبار الفرضيات

إضافة إلى تقدير المعاملات، تتضمن نتائج الانحدار الخطأ المعياري وإحصاءات التشخيص العادي، وسنطبق اختبار t على معاملات المتغيّر الوهمي، وفرضيتنا العدمية هي: $H_0:\delta=0$ والفرضية البديلة $H_1:\delta\neq 0$؛ وبعبارة أخرى تقول الفرضية العدمية أنه لا يوجد اختلاف بين التكاليف العامة لنوعي المدارس، وحيث احصائية t تساوي 6.40 سترفض الفرضية العدمية عند مستوى معنوية 0.1٪، ونستطيع تطبيق اختبار t على المعلمات الأخرى، واحصائية t لمعامل N تساوي 8.34؛ ومنها نستنتج أن الكلفة الحدية معنوية جداً وتختلف عن الصفر، وفي حالة المقطع احصائية t تساوي - 1.43 وعليه لا نستطيع رفض الفرضية العدمية $H_0:\beta_1=0$، وعليه تكون إحدى الشروحات حماقة الكلفة العامة السالبة للمدارس العادية، قد لا يكون لها أي تكاليف عامة وتقديرنا هو رقم عشوائي، والشكل الأكثر حقيقة لهذه الفرضية أن β_1 موجبة لكنها صغيرة وحد الخطأ مسؤول عن التقدير السالب، كما أن الاحتمال الإضافي هو أن النموذج هو سيء التوصيف.

صندوق (8-1): تفسير معاملات المتغيّر الوهمي في انحدار لوغاريتمي وشبه لوغاريتمي

افرض أنه يوجد لدينا نموذج الانحدار التالي:

$$\log Y = \beta_1 + \beta_2 \log X + \delta D + u$$

حيث D هو متغيّر وهمي و δ معامله، وسنعيد كتابة النموذج على النحو التالي:

$$\begin{aligned} Y &= e^{\beta_1 + \beta_2 \log X + \delta D + u} \\ &= e^{\beta_1} e^{\log X^{\beta_2}} e^{\delta D} e^{u} \\ &= e^{\beta_1} X^{\beta_2} e^{\delta D} e^{u} \end{aligned}$$

فيما يتعلق بالحد $e^{\delta D}$، فعندما تكون $D = 0$ يصبح الحد مساوياً e^0، ومن الطبيعي أن e^0 تساوي 1، وعليه سنضرب Y في e^0 وسيكون المتغيّر الوهمي ليس له أي أثر على الفئة المرجعية، أما عندما تكون $D = 1$ للفئة الأخرى يصبح الحد e^{δ} ونضرب Y في e^{δ}؛ فإذا كانت δ صغيرة ستساوي e^{δ} تقريباً $(1+\delta)$ تعني أن Y هي نسبة δ أكبر للفئة الأخرى من الفئة المرجعية، وإذا كانت δ ليست صغيرة سيكون الفرق النسبي $(e^{\delta} - 1)$.

النموذج شبه اللوغاريتمي:

$$\log Y = \beta_1 + \beta_2 X + \delta D + u$$

يمكن إعادة كتابتها:

$$Y = e^{\beta_1 + \beta_2 X + \delta D + u}$$
$$= e^{\beta_1} e^{\beta_2 X} e^{\delta D} e^{u}$$

أثر المتغيّر الوهمي وتفسير معامله يكون نفسه كما في النموذج اللوغاريتمي.

8-2- استخدام أكثر من متغيّر الوهمي

استخدمنا المتغيّر الوهمي في المبحث السابق للتفريق بين المدارس العادية والمهنية عند تقدير دالة التكاليف، وفي الحقيقية يوجد نوعين من المدارس العادية الثانوية في شنغهاي؛ هناك مدارس عادية فيها تعليم أكاديمي ومدارس مهنية، ومن اسمها تعني أن المدارس المعنية تعني نقل المهارات المهنية كما في التعليم الأكاديمي، وعلى كل حال، المكوّن المهني للمناهج هو صغير والمدارس المشابهة للمدارس العادية، أحياناً يوجد مدارس عامة يضاف إليها ورش عمل، وبالمثل يوجد نوعين من المدارس المهنية: مدارس تقنية تدرب الفنيين ومدارس العمال الماهرين تدرب الحرفيين.

سيكون للمتغيّر النوعي أربع فئات ونحتاج إلى تطوير مجموعة أكثر تفصيلاً للمتغيّرات الوهمية، والإجراء المعياري لاختيار إحدى الفئات فئة مرجعية للمعادلة الأساسية ثم نعرّف المتغيّرات الوهمية لكل فئة من الفئات الأخرى، بشكل عام تكون الممارسة الجيدة لاختيار الفئة السائدة أو الفئة الأكثر طبيعية، إذا كان أحدها فئة مرجعية. وفي مثال شنغهاي من المعقول لاختيار المدارس العامة الأكثر عدداً والمدارس الأخرى تختلف فيما بينها.

سنعرف المتغيّرات الوهمية للأنواع الثلاث الأخرى، $TECH$ متغيّر وهمي للمدارس التقنية: $TECH=1$ إذا كانت المشاهدة تتعلق بمدرسة تقنية و 0 لغير ذلك، وبالمثل نعرف المتغيّر الوهمي $WORKER$ و VOC لمدارس العمال الماهرين والمدارس المهنية، ويصبح نموذج الانحدار كما يلي:

$$COST = \beta_1 + \delta_T TECH + \delta_W WORKER + \delta_V VOC + \beta_2 N + u \qquad (8.8)$$

حيث δ_T و δ_W و δ_V هي معاملات تعرض الكلفة العامة الإضافية للتقني والعامل الماهر والمدارس المهنية بالنسبة لكلفة المدارس العامة.

يبين الجدول (8-2) بيانات أول 10 مدارس من 74 مدرسة، لاحظ كيف تحددت قيم المتغيّرات الوهمية $TECH$ و $WORKER$ و VOC حسب نمط المدرسة لكل مشاهدة.

جدول (8-2) النفقات المتكررة، وعدد من الطلاب، ونوع المدرسة

School	Type	*COST*	*N*	*TECH*	*WORKER*	*VOC*
1	Technical	345,000	623	1	0	0
2	Technical	537,000	653	1	0	0
3	General	170,000	400	0	0	0
4	Skilled Workers'	526	663	0	1	0
5	General	100,000	563	0	0	0
6	Vocational	28,000	236	0	0	1
7	Vocational	160,000	307	0	0	1
8	Technical	45,000	173	1	0	0
9	Technical	120,000	146	1	0	0
10	Skilled Workers'	61,000	99	0	1	0

وتم تقدير المعادلة وكانت نتائج الانحدار كما يلي (الخطأ المعياري بين الأقواس):

$$COST = -55{,}000 + 154{,}000\ TECH + 143{,}000\ WORKER$$
$$\quad\ (27000)\qquad\ (27000)\qquad\qquad (28000)$$

$$+ 53000\ VOC + 343\ N \qquad R^2 = 0.63$$
$$(31000)\qquad\quad (40)$$

Dependent Variable: COST
Method: Least Squares
Date: 08/19/15 Time: 12:42
Sample: 1 74
Included observations: 74

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-54893.09	26673.08	-2.057996	0.0434
TECH	154110.9	26760.41	5.758915	0.0000
WORKER	143362.4	27852.80	5.147144	0.0000
VOC	53228.64	31061.65	1.713645	0.0911
N	342.6335	40.21950	8.519090	0.0000

R-squared	0.632050	Mean dependent var	187418.0
Adjusted R-squared	0.610719	S.D. dependent var	141969.9
S.E. of regression	88578.37	Akaike info criterion	25.68634
Sum squared resid	**5.41E+11**	Schwarz criterion	25.84202
Log likelihood	-945.3946	Hannan-Quinn criter.	25.74844
F-statistic	29.63132	Durbin-Watson stat	1.330643
Prob(F-statistic)	0.000000		

يشير معامل N إلى أن الكلفة الحدية لكل طالب في السنة تساوي 343 يوان، كما ويشير معامل الحد الثابت إلى أن الكلفة العامة السنوية للمدارس الأكاديمية العامة تساوي - 55000 يوان في السنة؛ وهذا ليس له معنى ويشير إلى بعض الخطأ في النموذج، ومعامل $TECH$ و $WORKER$ و VOC يشير إلى أن الكلفة العامة للتقني والعامل الماهر والمدارس المهنية هي 154000 يوان و 143000 يوان و 53000 يوان أكبر من الكلفة العامة للمدارس.

8-2-1- مصيدة المتغيّر الوهمي

ماذا سيحدث إذا تم تضمين متغيّر وهمي للفئة المرجعية؟ هناك نتيجتين:

الأولى من الممكن حساب معاملات الانحدار، لكن لا تستطيع تقديم تفسير لها، المعامل b_1 هو تقدير للمقطع، ومعاملات المتغيّرات الوهمية هي تقديرات لزيادة المقطع من مستواه الأساسي، لكن لا يوجد تعريف للأساس وينهار التفسير.

النتيجة الأخرى هي أن الإجراء الرقمي لحساب معاملات الانحدار سوف تكسر ويرسل لك الكمبيوتر رسالة بالخطأ (من الممكن حذف أحد المتغيّرات الوهمية)، افرض وجود m فئة وهمية وعرفت المتغيّرات الوهمية $D_1, ..., D_m$ عند المشاهدة i يكون $\sum G_{ij} = 1$ لأن أحد المتغيّرات الوهمية يساوي 1 والبقية 0، لكن المقطع β_1 هو ناتج المعامل β_1 والمتغيّر الخاص هو القيمة 1 في المشاهدات، وبالتالي لجميع المشاهدات يساوي مجموع المتغيّرات الوهمية لهذا المتغيّر في نموذج الانحدار؛ وهذا يسمى **مصيدة المتغيّر الوهمي dummy variable trap**، وبالنتيجة النموذج مقيد بحالة خاصة للارتباط الذاتي التام، ويجعل من غير الممكن حساب معاملات الانحدار.

أما الإجراء البديل لتجنب مشكلة مصيدة المتغيّر الوهمي يكون بإسقاط المقطع من النموذج، وبإسقاط متغيّر الوحدة الخاصة تختفي العلاقة الخطية التامة بين المتغيّرات.

8-3- مَيل المتغيّر الوهمي

سنعود إلى مثال كلفة المدرسة، وفرضية الكلفة الحدية للطالب هي نفسها للمدارس المهنية والمدارس العادية وهذا غير واقعي؛ لأن المدارس المهنية تتكبد نفقات لمواد التدريب مرتبطة بعدد الطلاب، ونسبة المدرسين للطلاب هي أكبر في المدارس المهنية لأن مجموعات وورش العمل ليست أكبر من الصفوف الأكاديمية، ونستطيع تعديل الفرضية بإضافة ميل المتغيّر الوهمي $NOCC$ كناتج لضرب N و OCC:

$$COST = \beta_1 + \delta OCC + \beta_2 N + \lambda NOCC + u \qquad (8.9)$$

وبما أن ناتج المتغيّرين في التوصيف، فإن ميل المتغيّر الوهمي حالة خاصة في المتغيّر التفاعلي $(N \times OCC)$، ولأن أحد المتغيّرات في العملية التفاعلية متغيّر نوعي، يكون التفسير لميل المتغيّر الوهمي مباشرة أكثر من الحالة الخاصة، وفي المثال الحالي إذا أعيد كتابة (8.9):

$$COST = \beta_1 + \delta OCC + (\beta_2 + \lambda OCC)N + u \qquad (8.10)$$

تستطيع أن ترى هذا الأثر لميل المتغيّر الوهمي بالسماح لمعامل N للمدارس المهنية ليكون λ أكبر من المدارس العادية، فإذا كان OCC صفراً سيكون $NOCC$ صفراً وتصبح المعادلة كما يلي:

$$COST = \beta_1 + \beta_2 N + u \qquad (8.11)$$

إذا كان OCC يساوي واحداً تكون $NOCC$ تساوي N وتصبح المعادلة:

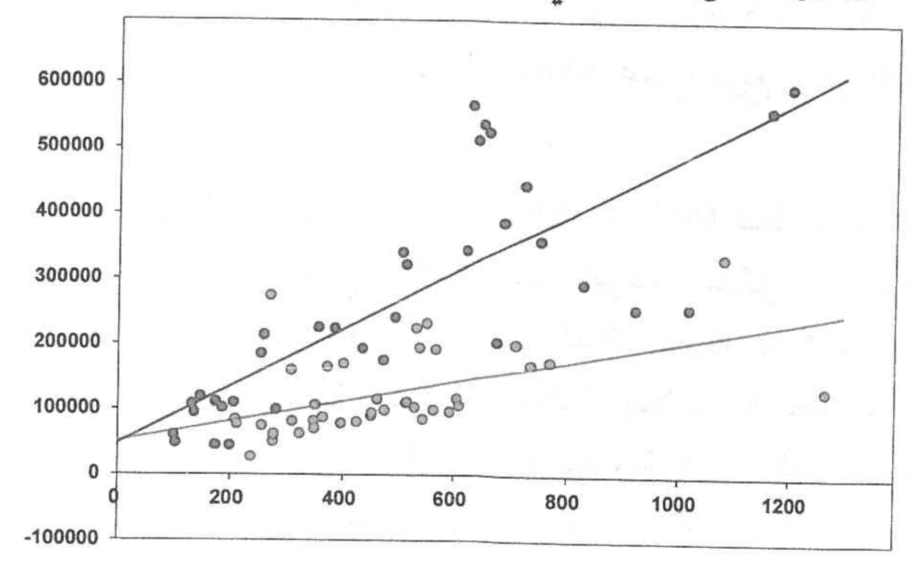
$$COST = \beta_1 + \delta + (\beta_2 + \lambda)N + u \qquad (8.12)$$

تصبح λ الكلفة الحدية الإضافية المرتبطة بالمدارس المهنية، وبنفس الطريقة تكون δ الكلفة الكلية الإضافية المرتبطة بهم، والجدول (8-3) يُظهر بيانات أول 10 مدارس في العينة:

جدول (8-2) النفقات المتكررة، وعدد من الطلاب، ونوع المدرسة

School	Type	*COST*	*N*	*OCC*	*NOCC*
1	Technical	345,000	623	1	623
2	Technical	537,000	653	1	653
3	General	170,000	400	0	0
4	Skilled Workers'	526	663	1	663
5	General	100,000	563	0	0
6	Vocational	28,000	236	0	0
7	Vocational	160,000	307	0	0
8	Technical	45,000	173	1	173
9	Technical	120,000	146	1	146
10	Skilled Workers'	61,000	99	1	99

ويظهر الشكل (8-3) دالتي التكاليف:



شكل رقم 8-3: دالة تكاليف المدارس العادية حسب ميل المتغيّر الوهمي

8-4- اختبار تشاو Chow test

في بعض الأحيان تتكون عينة المشاهدات من عينتين فرعيتين أو أكثر، ويكون من غير المؤكد تنفيذ انحدار واحد مشترك، أو انحدار منفصل لكل عينة فرعية، وفي الحقيقة التطبيق ليس مثل هذا تماماً، لأنه قد يكون في بعض المجالات مزيج لعينات فرعية تستخدم متغيّرات وهمية مناسبة وميل متغيّرات لتخفيف افتراض أن المعامل يجب أن يكون نفسه لكل عينة فرعية.

افرض أن لدينا عينة تتكون من عينتين فرعيتين وقد تتساءل هل امزجها في انحدار مجمّع pooled regression P أو تنفيذ انحدارين منفصلين A و B. سنشير إلى مجموع مربع بواقي انحدار العينة الفرعية SSR_A و SSR_B، ونشير إلى SSR_A^P و SSR_B^P إلى مجموع مربع بواقي الانحدار المجمّع للمشاهدات التي تخص العينتين الفرعيتين، وبما أن انحدار العينة الفرعية يخفّض SSR يكون قياسها بشكل عام أفضل من الانحدار التجميعي، وعليه يكون $SSR_A \le SSR_A^P$ و $SSR_B \le SSR_B^P$ و $(SSR_A + SSR_B) \le SSR_P$ حيث أن SSR_P مجموع مربع بواقي الانحدار المجمّع، ويساوي مجموع SSR_A^P و SSR_B^P.

تحدث المساواة بين SSR_P و $(SSR_A + SSR_B)$ فقط عندما تتطابق معاملات الانحدار المجمّع والانحدارات الفرعية. بشكل عام سنطور $(SSR_P - SSR_A - SSR_B)$ عندما تكون العينة منفصلة، هناك ثمن يدفع من k درجة حرية إضافية مستخدمة، حيث انه بدلاً من k معلمة للانحدار المجمّع يكون لدينا تقدير k لكل عينة فرعية وتكون $2k$ للجميع، وبعد فصل العينة لا يزال مجموع البواقي $(SSR_A + SSR_B)$ (غير المفسّر)، ويكون لدينا $n-2k$ درجة حرية.

نحن الآن في موقع لنرى التحسن في التقدير عندما نفصل العينة يكون معنوياً نجري اختبار F يسمى اختبار تشاو Chow test، وسنستخدم اختبار F التالي:

$$F(k, n-2k) = \frac{(SSR_P - SSR_A - SSR_B)/k}{(SSR_A + SSR_B)/(n-2k)} \qquad (8.13)$$

التي توزيعها k و $n-2k$ درجة حرية في ظل الفرضية العدمية لعدم معنوية تحسن التقدير.

Dependent Variable: COST
Method: Least Squares
Date: 08/19/15 Time: 19:32
Sample: 1 74 **IF OCC=0**
Included observations: 40

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	51475.25	21599.14	2.383208	0.0223
N	152.2982	41.39782	3.678896	0.0007

R-squared	0.262627	Mean dependent var	123809.3
Adjusted R-squared	0.243222	S.D. dependent var	64999.34
S.E. of regression	56544.87	Akaike info criterion	24.77216
Sum squared resid	**1.21E+11**	Schwarz criterion	24.85661
Log likelihood	-493.4433	Hannan-Quinn criter.	24.80270
F-statistic	13.53427	Durbin-Watson stat	2.071392
Prob(F-statistic)	0.000723		

سنشرح اختبار Chow بالاعتماد على بيانات دالة تكاليف المدرسة بوضع تمييز بين المدارس العادية والمدارس المهنية، وسنحتاج إلى تنفيذ ثلاثة انحدارات: الأول انحدار $COST$ على N باستخدام العينة كاملة 74 مدرسة،

وهذا ما قدرناه في المبحث (8-1) وهو الانحدار المجمّع؛ حيث أن *SSR* له تساوي 8.92×10^{11}. وفي الانحدار الثاني والثالث ننفذ نفس الانحدار للعينتين الفرعيتين للمدارس العادية والمدارس المهنية كل على حده، آخذين بالاعتبار *SSR* لكل منهما، ونتائج انحدارات العينات الفرعية تبينها النتائج أدناه وخط الانحدار.

Dependent Variable: COST
Method: Least Squares
Date: 08/19/15 Time: 19:41
Sample: 1 74 **IF OCC=1**
Included observations: 34

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	47974.07	33879.03	1.416040	0.1664
N	436.7769	58.62085	7.450879	0.0000

R-squared	0.634351	Mean dependent var	262251.7
Adjusted R-squared	0.622924	S.D. dependent var	170056.2
S.E. of regression	104425.5	Akaike info criterion	26.00736
Sum squared resid	**3.49E+11**	Schwarz criterion	26.09714
Log likelihood	-440.1251	Hannan-Quinn criter.	26.03798
F-statistic	55.51561	Durbin-Watson stat	1.074550
Prob(F-statistic)	0.000000		

حصلنا على *SSR* للمدارس العادية 1.21×10^{11}، و 3.49×10^{11} للمدارس المهنية، ومجموع *SSR* لانحدار العينات الفرعية يساوي 4.70×10^{11}، ويجب أن تكون أقل من *SSR* للانحدار المجمّع، ولنرى فيما إذا كانت أقل معنوية سننفذ اختبار Chow، حيث بسط إحصائية F يساوي $(8.92 - 4.70) \times 10^{11}$ مقسوماً على الكلفة حسب مفهوم درجات الحرية، وهي تساوي اثنين لأننا نقدر مقطعين ومعاملي ميل اثنين بدلاً من 1 لكل

منهما، والمقام يساوي SSR المشترك بعد فصل العينة 4.70×10^{11} مقسوماً على درجات الحرية المشتركة والتي تساوي 70 حيث لدينا 74 مشاهدة ولدينا 4 درجات حرية من تقدير 2 معلمة في كل معادلة، وعندما نحسب إحصائية F نحذف 10^{11} يكون لدينا:

$$F(2,70)=\frac{(8.92-4.70)\times 10^{11}/2}{(4.70\times 10^{11})/70}=31.4 \qquad (8.14)$$

والقيمة الحرجة للإحصائية $F(2,70)$ عند مستوى معنوية 0.1٪ هي 7.64، وعليه نحصل على نتيجة تكون معنوية لتقدير عينات منفصلة، وعليه سوف لا نستخدم انحدار لبيانات مجمّعة إنما نقدر انحدارين منفصلين.

تطبيق عملي

لاختبار استقرار السلسلة الزمنية المستخدمة في تحليل دالة الإنتاج في قطاع الصناعة التحويلية في الأردن خلال الفترة 1971-2005 حسب دالة إنتاج كوب-دوغلاس التالية:

$$Q_t=\beta_0+\beta_1 L_t+\beta_2 K_t+\varepsilon_t$$

حيث أن: Q_t اللوغاريتم الطبيعي لإجمالي الصناعة التحويلية في الأردن للفترة t

L_t اللوغاريتم الطبيعي لعنصر العمل في الصناعة التحويلية في الأردن للفترة t

K_t اللوغاريتم الطبيعي لعنصر رأس المال في الصناعة التحويلية في الأردن للفترة t

ε_t حد الخطأ العشوائي

لاختبار تكافؤ معلمات الانحدار بين مجموعتي البيانات؛ لفترة تسبق عام 1989 (1971-1988) وفترة من عام 1989 وما بعدها (1991-2005)، فيما إذا كانتا تحتويان معلمات انحدار معنوية لنفس المعادلة النظرية، وهذا يساعدنا أن نقرر فيما إذا كان من المناسب دمج مجموعتي البيانات في سلسلة واحدة، وتكون الفرضية الأساسية أن معلمات الميل متكافئة في العينتين، وقد نستخدم المتغيّرات الوهمية للتمييز بين مجموعة البيانات، وإذا أردنا تطبيق اختبار Chow test للقطع الهيكلي نتبع ما يلي:

1- نقدر انحداري الاختبار للعينة الأولى والعينة الثانية ونستخرج قيم SSR للعينتين وكانتا على النحو التالي: $SSR_{n_1} = 0.113689$ و $SSR_{n_2} = 0.074615$.

Dependent Variable: LOG(OUTPUT)
Method: Least Squares
Included observations: 10

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.014421	2.575954	-0.393804	0.7054
LOG(CAPITAL)	1.006927	0.217624	4.626900	0.0024
LOG(LABOR)	-0.475651	0.411576	-1.155683	0.2857

R-squared	0.909972	Mean dependent var	6.428522
Adjusted R-squared	0.884250	S.D. dependent var	0.374584
S.E. of regression	0.127441	Akaike info criterion	-1.038994

Sum squared resid	**0.113689**	Schwarz criterion	-0.948218
Log likelihood	8.194970	Hannan-Quinn criter.	-1.138575
F-statistic	35.37680	Durbin-Watson stat	2.286267
Prob(F-statistic)	0.000219		

Dependent Variable: LOG(OUTPUT)
Method: Least Squares
Included observations: 17

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	11.38210	1.080558	10.53353	0.0000
LOG(CAPITAL)	-0.106149	0.111820	-0.949291	0.3586
LOG(LABOR)	1.255376	0.136303	9.210199	0.0000

R-squared	0.976559	Mean dependent var	7.800770
Adjusted R-squared	0.973210	S.D. dependent var	0.446028
S.E. of regression	0.073004	Akaike info criterion	-2.237808
Sum squared resid	**0.074615**	Schwarz criterion	-2.090770
Log likelihood	22.02137	Hannan-Quinn criter.	-2.223192
F-statistic	291.6178	Durbin-Watson stat	1.347535
Prob(F-statistic)	0.000000		

2- يتم تجميع بيانات كلا العينتين ونقدر الانحدار المبين أعلاه للسلسلة كاملة 1971-2005، ونحصل على $SSR_n = 0.477714$.

Dependent Variable: LOG(OUTPUT)
Method: Least Squares
Included observations: 27

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.493853	1.415059	5.295788	0.0000
LOG(CAPITAL)	0.318951	0.145950	2.185344	0.0389
LOG(LABOR)	0.864833	0.174654	4.951703	0.0000

R-squared	0.970696	Mean dependent var	7.292530
Adjusted R-squared	0.968254	S.D. dependent var	0.791838
S.E. of regression	0.141084	Akaike info criterion	-0.974481
Sum squared resid	**0.477714**	Schwarz criterion	-0.830499
Log likelihood	16.15550	Hannan-Quinn criter.	-0.931668
F-statistic	397.5056	Durbin-Watson stat	0.768514
Prob(F-statistic)	0.000000		

3- نحسب احصائية F كما يلي:

$$F = \frac{(SSR_n - [SSR_{n_1} + SSR_{n_2}])/k}{(SSR_{n_1} + SSR_{n_2})/(n_1 + n_2 - 2k)}$$

$$= \frac{(0.477714 - [0.113689 + 0.074615])/3}{(0.113689 + 0.074615)/(10 + 17 - 2 \times 3)}$$

$$= \frac{0.28941/3}{0.188304/21} = 10.758$$

4- بما أن:

$$F_{\text{المحسوبة}} = 10.758 > F_{\text{الجدولية}} = 3.07$$

سنرفض الفرضية الأساسية، وهذا يعني أن معلمات كلا الانحدارين غير متكافئة، أي أن السلسلة الزمنية غير مستقرة ونستنتج عدم وجود دليل على الاستقرار الهيكلي، وبالتالي لا نستطيع تقدير المعادلة بكامل بيانات العينة، ويوجد عملية قطع للبيانات عند 1989، ويتم تجزئة العينة إلى عينتين

ونقدر معادلة لبيانات الفترة الأولى 1971-1988، ومعادلة لبيانات الفترة الثانية 1991-2005.

تمارين

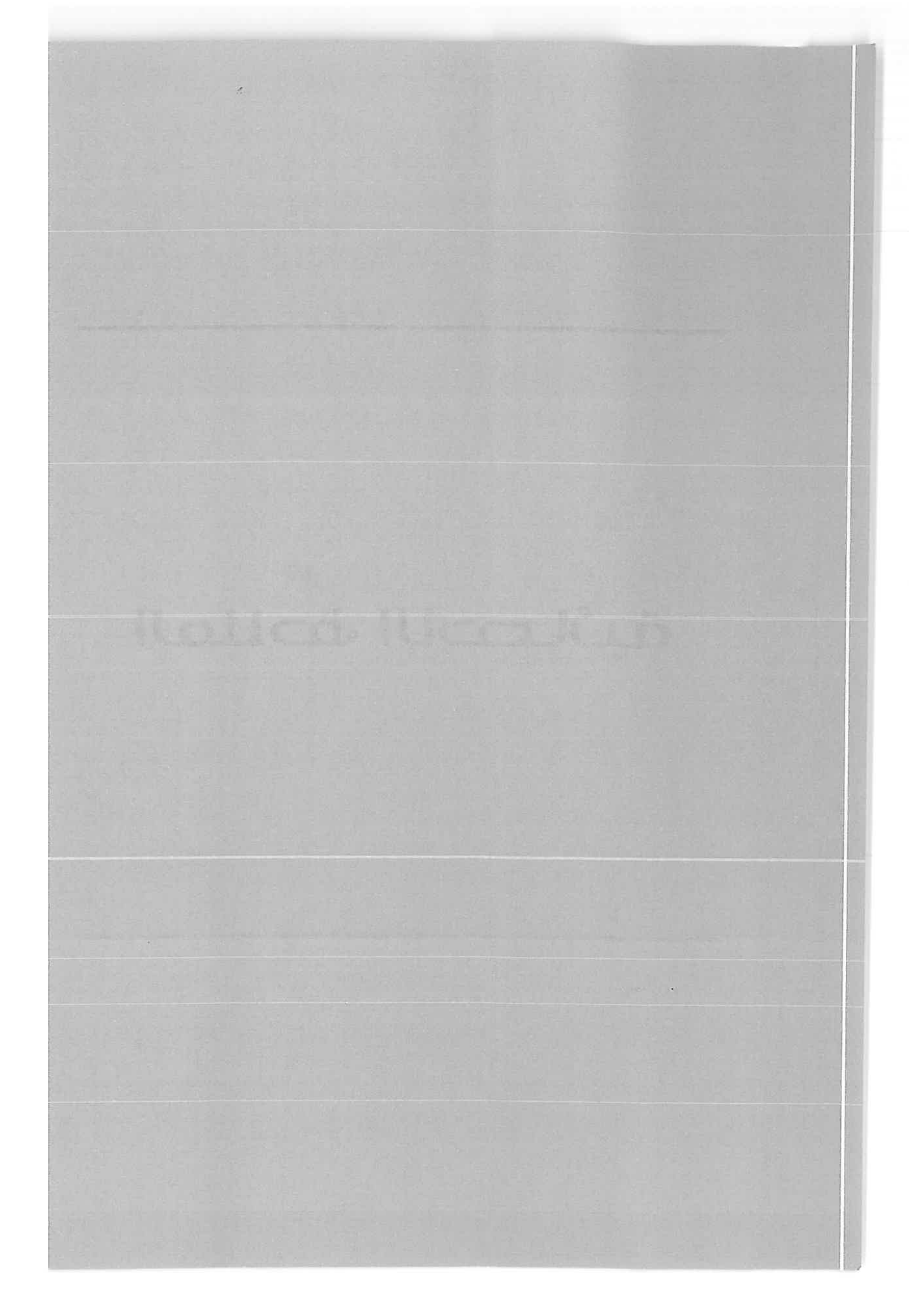
8-1- اشرح كيفية استخدام المتغيّرات الوهمية لمعلومات نوعية كمية في نموذج الانحدار مستخدماً مثالاً من النظرية الاقتصادية.

8-2- بيّن أثر استخدام متغيّر وهمي ثنائي على الحد الثابت وعلى ميل معامل الميل في نموذج انحدار بسيط.

8-3- أعطي مثالاً من النظرية الاقتصادية تستخدم فيه متغيّراً وهمياً موسمياً، واشرح لماذا لا نستخدم جميع المتغيّرات الوهمية مع بعضها في نفس المعادلة عندما تحتوي على الحد الثابت، ويجب استثناء أحدها الذي سيتصرف كمتغيّر وهمي مرجعي. وماذا نعني بالمتغيّر الوهمي المرجعي؟

8-4- صف الخطوات التي تتبع عند إجراء اختبار Chow للاستقرار الهيكلي.

الملاحق الاحصائية



جدول رقم (1)

Critical Values of the *t* Distribution

	Significance Level				
1-Tailed: 2-Tailed:	.10 .20	.0 5 .1	.0 25 .05	.0 1 .0	.005 .010
Degrees of Freedom					
1	3.078	6.314	12.7	31.821	63.657
2	1.886	2.920	06	6.965	9.925
3	1.638	2.353	4.3	4.541	5.841
4	1.533	2.132	03	3.747	4.604
5	1.476	2.015	3.1	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
40	1.303	1.684	2.021	2.423	2.704
60	1.296	1.671	2.000	2.390	2.660
90	1.291	1.662	1.987	2.368	2.632
120	1.289	1.658	1.980	2.358	2.617
∞	1.282	1.645	1.960	2.326	2.576

Examples: The 1% critical value for a one-tailed test with 25 *df* is 2.485. The 5% critical for a two-tailed test with large (> 120) *df* is 1.96.
Source: This table was generated using the Stata® function invt.

٩٩٪

جدول رقم (2)

1% Critical Values of the F Distribution

Denominator Degrees of Freedom	Numerator Degrees of Freedom									
	1	2	3	4	5	6	7	8	9	10
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63
90	6.93	4.85	4.01	3.54	3.23	3.01	2.84	2.72	2.61	2.52
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32

Example: The 1% critical value for numerator $df = 3$ and denominator $df = 60$ is 4.13.
Source: This table was generated using the Stata® function invfprob.

5% Critical Values of the F Distribution

Denominator Degrees of Freedom	Numerator Degrees of Freedom: 1	2	3	4	5	6	7	8	9	10
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99
90	3.95	3.10	2.71	2.47	2.32	2.20	2.11	2.04	1.99	1.94
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83

Example: The 5% critical value for numerator $df = 4$ and large denominator $df\ (\infty)$ is 2.37.
Source: This table was generated using the Stata® function invfprob.

10% Critical Values of the F Distribution

Denominator Degrees of Freedom	Numerator Degrees of Freedom: 1	2	3	4	5	6	7	8	9	10
10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32
11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25
12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19
13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14
14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06
16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03
17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00
18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98
19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98	1.96
20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94
21	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95	1.92
22	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90
23	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92	1.89
24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88
25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89	1.87
26	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86
27	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87	1.85
28	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84
29	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86	1.83
30	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82
40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76
60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71
90	2.76	2.36	2.15	2.01	1.91	1.84	1.78	1.74	1.70	1.67
120	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68	1.65
∞	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63	1.60

Example: The 10% critical value for numerator $df = 2$ and denominator $df = 40$ is 2.44.
Source: This table was generated using the Stata® function invfprob.

جدول رقم (3)

Critical Values of the Chi-Square Distribution

Degrees of Freedom	Significance Level .10	.05	.01
1	2.71	3.84	6.63
2	4.61	5.99	9.21
3	6.25	7.81	11.34
4	7.78	9.49	13.28
5	9.24	11.07	15.09
6	10.64	12.59	16.81
7	12.02	14.07	18.48
8	13.36	15.51	20.09
9	14.68	16.92	21.67
10	15.99	18.31	23.21
11	17.28	19.68	24.72
12	18.55	21.03	26.22
13	19.81	22.36	27.69
14	21.06	23.68	29.14
15	22.31	25.00	30.58
16	23.54	26.30	32.00
17	24.77	27.59	33.41
18	25.99	28.87	34.81
19	27.20	30.14	36.19
20	28.41	31.41	37.57
21	29.62	32.67	38.93
22	30.81	33.92	40.29
23	32.01	35.17	41.64
24	33.20	36.42	42.98
25	34.38	37.65	44.31
26	35.56	38.89	45.64
27	36.74	40.11	46.96
28	37.92	41.34	48.28
29	39.09	42.56	49.59
30	40.26	43.77	50.89

Example: The 5% critical value with $df = 8$ is 15.51.
Source: This table was generated using the Stata® function invchi

جدول رقم (4)

Lower and upper 1% critical values for Durbin–Watson statistic

	$k'=1$		$k'=2$		$k'=3$		$k'=4$		$k'=5$	
T	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
15	0.81	1.07	0.70	1.25	0.59	1.46	0.49	1.70	0.39	1.96
16	0.84	1.09	0.74	1.25	0.63	1.44	0.53	1.66	0.44	1.90
17	0.87	1.10	0.77	1.25	0.67	1.43	0.57	1.63	0.48	1.85
18	0.90	1.12	0.80	1.26	0.71	1.42	0.61	1.60	0.52	1.80
19	0.93	1.13	0.83	1.26	0.74	1.41	0.65	1.58	0.56	1.77
20	0.95	1.15	0.86	1.27	0.77	1.41	0.68	1.57	0.60	1.74
21	0.97	1.16	0.89	1.27	0.80	1.41	0.72	1.55	0.63	1.71
22	1.00	1.17	0.91	1.28	0.83	1.40	0.75	1.54	0.66	1.69
23	1.02	1.19	0.94	1.29	0.86	1.40	0.77	1.53	0.70	1.67
24	1.04	1.20	0.96	1.30	0.88	1.41	0.80	1.53	0.72	1.66
25	1.05	1.21	0.98	1.30	0.90	1.41	0.83	1.52	0.75	1.65
26	1.07	1.22	1.00	1.31	0.93	1.41	0.85	1.52	0.78	1.64
27	1.09	1.23	1.02	1.32	0.95	1.41	0.88	1.51	0.81	1.63
28	1.10	1.24	1.04	1.32	0.97	1.41	0.90	1.51	0.83	1.62
29	1.12	1.25	1.05	1.33	0.99	1.42	0.92	1.51	0.85	1.61
30	1.13	1.26	1.07	1.34	1.01	1.42	0.94	1.51	0.88	1.61
31	1.15	1.27	1.08	1.34	1.02	1.42	0.96	1.51	0.90	1.60
32	1.16	1.28	1.10	1.35	1.04	1.43	0.98	1.51	0.92	1.60
33	1.17	1.29	1.11	1.36	1.05	1.43	1.00	1.51	0.94	1.59
34	1.18	1.30	1.13	1.36	1.07	1.43	1.01	1.51	0.95	1.59
35	1.19	1.31	1.14	1.37	1.08	1.44	1.03	1.51	0.97	1.59
36	1.21	1.32	1.15	1.38	1.10	1.44	1.04	1.51	0.99	1.59
37	1.22	1.32	1.16	1.38	1.11	1.45	1.06	1.51	1.00	1.59
38	1.23	1.33	1.18	1.39	1.12	1.45	1.07	1.52	1.02	1.58
39	1.24	1.34	1.19	1.39	1.14	1.45	1.09	1.52	1.03	1.58
40	1.25	1.34	1.20	1.40	1.15	1.46	1.10	1.52	1.05	1.58
45	1.29	1.38	1.24	1.42	1.20	1.48	1.16	1.53	1.11	1.58
50	1.32	1.40	1.28	1.45	1.24	1.49	1.20	1.54	1.16	1.59
55	1.36	1.43	1.32	1.47	1.28	1.51	1.25	1.55	1.21	1.59
60	1.38	1.45	1.35	1.48	1.32	1.52	1.28	1.56	1.25	1.60
65	1.41	1.47	1.38	1.50	1.35	1.53	1.31	1.57	1.28	1.61
70	1.43	1.49	1.40	1.52	1.37	1.55	1.34	1.58	1.31	1.61
75	1.45	1.50	1.42	1.53	1.39	1.56	1.37	1.59	1.34	1.62
80	1.47	1.52	1.44	1.54	1.42	1.57	1.39	1.60	1.36	1.62
85	1.48	1.53	1.46	1.55	1.43	1.58	1.41	1.60	1.39	1.63
90	1.50	1.54	1.47	1.56	1.45	1.59	1.43	1.61	1.41	1.64
95	1.51	1.55	1.49	1.57	1.47	1.60	1.45	1.62	1.42	1.64
100	1.52	1.56	1.50	1.58	1.48	1.60	1.46	1.63	1.44	1.65

Note: T, number of observations; k', number of explanatory variables (excluding a constant term).

Source: Durbin, J. and Watson, G.S. (1951) Testing for serial correlation in least squares regression II *Biometrika*, 38(1–2), 159–177. Reprinted with the permission of Oxford University Press.

المراجع

- السواعي، خالد محمد، **الاقتصاد القياسي: المبادئ الأساسية وحالات تطبيقية**، 2018، دار الكتاب الثقافي، اربد – الأردن.
- السواعي، خالد محمد، **مدخل إلى القياس الاقتصادي**، 2015، الدار العربية للعلوم ناشرون، بيروت، لبنان.
- السواعي، خالد محمد، **أساسيات القياس الاقتصادي باستخدام *EViews***، 2012، دار الكتاب الثقافي، اربد-الأردن.

- Asteriou, Dimitrios and Hall, Stephen G., 2007. *Applied Econometrics: A Modern Approach*, Palgrave, revised edition.
- Brooks, Chris, 2006. *Introductory Econometrics for Finance*, Cambridge University Press, 7th edition.
- Dongherty, Christopher, 2011, *Introduction to Econometrics*, Oxford University Press, 4th edition.
- Griffiths, William E.; and Hill, R. Carter, 1993. *Learning and Practicing Econometrics,* John Wiley and Sons Inc.
- Hill, R. Carter; Griffiths, William E.; and Lim, Guay C., 2008. *Principles of Econometrics,* Wiley; 3nd edition.
- Koop, Gary, 2008. *Introduction to Econometrics*, Wiley.
- Studenmund, A. H., 2006. *Using Econometrics: A Practical Guide*, Addison Wesly, 5th edition.
- Thomas, R. L., 1997. *Econometrics: an introduction*, Prentics Hall.
- Vogelvang, Marno, 2005. *Econometrics: Theory and Application with EViews*, Prentice Hall.

د. خالد محمد السواعي
أستاذ الاقتصاد المساعد
جامعة الزرقاء
+962-7-9527-9666
khsawaie@yahoo.com

صدر للمؤلف

1. **الاقتصاد القياسي: المبادئ الأساسية وحالات تطبيقية**، دار الكتاب الثقافي، 2018. اربد – الأردن.
2. **مبادئ الاقتصاد القياسي**، دار الكتاب الثقافي، 2018. اربد – الأردن.
3. **مدخل إلى القياس الاقتصادي**، 2015، الدار العربية للعلوم ناشرون، بيروت- لبنان.
4. **موضوعات متقدمة في القياس الاقتصادي**، 2015، الدار العربية للعلوم ناشرون، بيروت- لبنان.
5. **التجارة والتنمية**، 2014، دار المناهج، الطبعة الثانية، عمان- الأردن.
6. **EViews والقياس الاقتصادي**، 2012، دار الكتاب الثقافي، اربد- الأردن.
7. **أساسيات القياس الاقتصادي باستخدام EViews**، 2012، دار الكتاب الثقافي، اربد- الأردن.
8. **مدخل إلى تحليل البيانات باستخدام SPSS**، 2011، عالم الكتب الحديث، اربد- الأردن.
9. **التجارة الدولية: النظرية وتطبيقاتها**، 2010، عالم الكتب الحديث، اربد- الأردن.
10. **دليل الإجراءات الجمركية**، 2000، دائرة الجمارك، عمان- الأردن.